Report on the ENUMERATE Thematic Surveys on Digital Collections in European Cultural Heritage Institutions 2013

by

Natasha Stroeker and René Vogels, Panteia (NL)
Gerhard Jan Nauta and Marco de Niet, DEN (NL)
on behalf of the ENUMERATE Thematic Network

September 2013

EUMERATE is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Contents

1 EXECUTIVE SUMMARY 4

2 INTRODUCTION 7

3 RESPONSE TO THE SURVEY 9

4 ACCESS 12

5 SIZE 19

5.1 Collection registration system (CRS) approach 20

5.2 Detailed measuring of digital sub-collections approach 22

5.3 Improved ENUMERATE Core Survey estimation approach 24

5.4 Evaluative remarks of the respondents 26

6 COST 28

6.1 COSTS IN THE ANNUAL REPORTS OF MEMORY ORGANISATIONS 29

6.2 THE COST OF SPECIFIC DIGITAL HERITAGE COLLECTIONS 30

6.3 CONCLUSIONS 37

7 IMPROVEMENTS FOR CORE SURVEY 2 AND OVERALL CONCLUSIONS 38

APPENDIX 1 REPORT MEETING DIGITAL PRESERVATION (17-01-2013) 41

APPENDIX 2 QUESTIONNAIRES: ACCESS, SIZE AND COST 43

APPENDIX 3 PREFERRED AND MOST ACCURATE APPROACH - SIZE 84

1 Executive summary

This report is the result of Task 3.4: *Implementation of Thematic Survey* of the ENUMERATE Thematic Network. This work is the second of three efforts into the state of digitisation in EU member states. The first was a 'general' Core Survey 1 in 2012[1], and this will be followed by another similar Core Survey 2 in late 2013.

The aims of Thematic Survey were: to test some of questions asked in Core Survey 1 in order to improve them in Core Survey 2; and to show the qualitative results of a set of additional questions not asked in Core Survey 1.

The work was formulated by a series of workshops, throughout Europe, in 2012 on the topics:

- *Size and Growth of the Digital Heritage Collection*;
- *Cost of Digital Cultural Heritage*;
- *Measuring the Use and Impact of Digital Cultural Heritage*;
- *Monitoring and Measuring Digital Preservation Practices*.

These workshops, and other consultations, concluded that there should be three separate surveys on Access Size, and Cost[2]. The nature of these surveys varied in style and implementation. Implementation of the surveys took place in the first half of 2013.

The Thematic Survey had the following outcomes:

ACCESS

Suggested changes for Core Survey 2:

- For the question '*Percentage of digital objects accessible*' add a new category – '*social media platforms like Flickr or Facebook*'.
- For the question '*Methods of use to measure the number of times digital metadata and/or digital objects are being accessed*' add question – '*When using web statistics, which are used? (Google Analytics or other?)*'

SIZE

Interesting outcomes:

- Even though the sample is quite small, the dataset gathered seems to confirm most of the outcomes from the ENUMERATE Core Survey 1, both about the size of the digital collections and the popularity of specific collections types for digitisation.
- Progress has been made in better understanding the composition of born digital heritage collections across Europe.

Suggested changes for Core Survey 2:

- The 'Collections Registration System (CRS)'[3] approach was the most time effective for the respondents. However the output from this approach was not as 'rich' as in the

[1] See: **Stroeker, Natasha and Vogels, René**. *Survey Report on Digitisation in European Cultural Heritage Institutions 2012*. ENUMERATE Thematic Network. May 2012.
Downloadable from: **http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2012.pdf**
[2] A survey on digital preservation was rejected. See page 5 for discussion.
[3] In some countries these systems are called Collection Management Systems (CMS).

other two approaches ('Detailed measuring of digital sub-collections', and 'Improved ENUMERATE Core Survey estimation.).

- There was little discussion about the classification that ENUMERATE created to designate digital heritage collection types, so this improved classification can be used in future surveys.
- Use the term 'descriptive metadata' instead of 'metadata'.

COST

Interesting outcomes:

- A minority of institutions report separately on their digital activities in their annual reports. If this does not change, it seems unlikely that future surveys will be able to collect really big datasets on costs of digitisation.
- Although the sample is small, there seems to be a trend that the cultural heritage institutions in Europe tend to spend more incidental costs on outsourcing activities than structural costs. Long term management of digital collections is very much an in-house activity.
- There is no one way that institutions record their costs. Some institutions use activity based costing, while others use another method.
- Incidental costs could be identified, while structural costs were much more difficult to identify.
- Institutions use a large variety of different categories of costs. Some new categories mentioned are activity based while others are not activity based.
- We understand better now which cost types can be more easily assessed by some institutions than others.

Suggested change for Core Survey 2:

- We will make adjustments to the question on expenditure. In the Thematic Survey we had defined a list of both structural and incidental cost types. Even though some institutions seemed to have some problems with this distinction, due to lack of a harmonised vocabulary, we will continue to use these cost types in the Core Survey 2 and ask for expenditures. The next question will be on making a division of the total expenditures in structural and incidental costs. This can be an estimate in percentages.

The report also has some general conclusions:

As the Thematic Survey was intended to test and improve the methodology with a smaller sample of institutions, there are few general conclusions that can be drawn about the status of digital cultural heritage in Europe. Most findings will feed into the preparations of the 2nd ENUMERATE Core Survey. The most important conclusions for this are:

- Although the questionnaires were rather complex and time consuming they give us useful input for Corey Survey 2 and beyond. On average they took a lot of time to fill out, even from institutions that were very motivated to participate in the Thematic Survey. There were some that did not fill out the questionnaire in the end although they had said they would do so. So careful consideration is needed to find the balance between on the

one hand a questionnaire that can provide rich data to better understand the progress of digital cultural heritage in Europe, and on the other hand an approach that does not scare away institutions from participating because of complexity or because participation is time-consuming.

- It is clear that these other costs are on average either in-house costs or costs of outsourcing. The overall percentages (In house 72%; outsourced 28%) are close to the percentages in the table of pre-defined costs (In house 74%. outsourced 26%).

Overall: the Thematic Survey gave us insight into what we should not ask, what we can ask in a different way and what we can change to the Core Survey 2. Also the outcomes give input to further discussions within the ENUMERATE team in the area of measuring the status of digital cultural heritage in Europe.

2 Introduction

The ENUMERATE Thematic Survey was designed to further develop the methodology to assess the status of digital cultural heritage in Europe. The Thematic Survey followed the ENUMERATE Core Survey 1 (2011/2012), and addressed certain questions raised about the methodology or the findings resulting from Core Survey 1. The results of the Thematic Survey, which was actually a set of sub-surveys, are presented in this report and will feed into the preparations of the ENUMERATE Core Survey 2 (2013/2014).

The topics of the Thematic Survey and research methods

The Thematic Survey consists of four parts, addressing the topics that are an integral part of the ENUMERATE methodology: Size, Cost, Access and Digital Preservation. The contents of the four parts of the Thematic Survey were prepared in four specialist meetings[4]:

- *Size and Growth of the Digital Heritage Collection,* January 2012, The Hague, Netherlands;
- *Cost of Digital Cultural Heritage*, March 2012, London, United Kingdom;
- *Measuring the Use and Impact of Digital Cultural Heritage*, June 2012, Madrid, Spain;
- *Monitoring and Measuring Digital Preservation Practices*, October 2012, The Hague, Netherlands.

For the topics of Size, Cost and Access web questionnaires were used. The Cost questionnaire consisted of two parts, sent out in two stages, and referred to as 'Cost A' and 'Cost B'.

For Digital Preservation an extra specialist meeting was held as part of the Thematic Survey, following the recommendations from the specialist meeting on this topic in October 2012. A separate high level questionnaire on digital preservation as part of the Thematic Survey was considered to be too complex, too premature and as a result potentially ineffective. The report on this extra meeting is included as Appendix 1 of this report.

Aim and deliverables Thematic Surveys on Access, Size and Cost

The aims of the Thematic Surveys on Access, Size and Cost were:

- Testing some questions in Core Survey 1 in order to improve them in Core Survey 2;
- Showing the qualitative outcomes of some additional questions (not in Core Survey 1).

An online questionnaire was chosen as the main instrument in doing research on the topics. In this way the team was able to make consistent use of knowledge and tools created in the context of ENUMERATE. There were also the pragmatic reasons of cost and time for using an internet questionnaire for all three sub-studies.

Since the questionnaires aimed to improve the methodology of the specific topic, some questions had an accompanying section for evaluative remarks. Respondents were encouraged make any comments that they thought might be useful for the development of a measuring methodology.

The result of the work were:

[4] http://www.enumerate.eu/fileadmin/ENUMERATE/documents/All_ENUMERATE_Specialist_Meeting.pdf

- This report, presenting the main results;
- Additional qualitative information;
- Input for adjusting questions in Core Survey 2. Only a limited number of changes were made in order to preserve the comparability of results with Core Survey 1.

Preparations and participating countries

Preparations for the Thematic Survey (TS) questionnaires started in September 2012. In November an announcement was sent to the national coordinators explaining what the ENUMERATE Team was trying to achieve, and asking the national coordinators to state whether they wanted to participate in the Thematic Survey, and in which part of the Thematic Survey they preferred to be involved. Also the need to translate the separate questionnaires was discussed. For most countries an English version was considered to be sufficient. If a translation was needed, it was provided by the national coordinators themselves (i.e. in Portugal, Slovakia and Poland). In mid-December, an inventory of all answers was made and the national coordinators were informed about which part of the TS that was selected for them. Early January, another request was sent out in order to provide the ENUMERATE Team with the names of institutions and contact persons. It took until April to collect all the contact details needed. The countries participating in the Thematic Survey are:

- COST A: Cyprus, Denmark and the Netherlands;
- ACCESS: Hungary, Spain, Portugal, Slovakia and Sweden;
- SIZE: Austria, Slovenia, Estonia, Finland, Germany, Ireland and Luxembourg;
- COST B: Belgium, UK, Switzerland, Lithuania, Greece, Poland, the Netherlands, Czech Republic.

Countries that were invited but did *not* participate in the Thematic Survey were: Bulgaria, France, Italy, Latvia, Romania and Malta. Malta was prepared to join *COST A* but did not manage to deliver the contact details of institutions to be invited.

3 Response to the survey

COST A was a detailed questionnaire, sent out to a small group of institutions. After a month the ENUMERATE Team had to conclude that the response on the COST A questionnaire was too low. To encourage more institutions to participate, a simplified version was sent out at the beginning of April (COST B). This improvement led to better responses. In this report, the responses from COST A are left out, only the responses to COST B are analysed in more detail.

For the SIZE questionnaire it was decided to use three methods of gathering data. This was because in Core Survey 1 an optional question was added to acquire more detailed information about collection size and growth, but only 10% of all respondents responded to this question. The methods were:

- Measuring collection size based on data from the collection registration system (collections management system) (A1);
- Detailed measuring of digital sub-collections (A2);
- Improved ENUMERATE Core Survey questions (A3).

Because the ENUMERATE team did not want to make the efforts required from the institutions too onerous, it was decided to give each institution only the questions related to two of the three methods. These two were randomly divided, so at the end each method was answered as often as the other one, i.e. A1 was answered by 33 institutions, A2 by 34 institutions and A3 by 34 institutions.

The ACCESS questionnaire had 22 questions and concerned the access and use of digital collections. Topics were: online access, online user services, online user activities, and licensing and legal issues.

The process and response to the Thematic Survey is summarised in the table below:

Table 3.1: Process and response of the Thematic Survey

Part of Thematic Survey	Access	SIZE	Cost B
Participating countries (number of invited institutions in brackets)	Hungary (5) Spain (9) Portugal (10) Slovakia (5) Sweden (7)	Austria (14) Slovenia (15) Estonia (13) Finland (13) Germany (11) Ireland (12) Luxembourg (25)	Belgium (11) UK (8) Switzerland (5) Lithuania (10) Greece (42) Poland (26) Netherlands (29) Czech Republic (50)
Starting date	March 8th	March 15th	April 3rd April 24th (Czech)
Final date	April 30th	April 30th	May 30th

Numbers:			
- target	30	70-100	70-100
- invited	36	103	181
- responded	**24**	**55**	**35**
Response (%)	**64%**	**53%**	**19%**

Table 3.1 shows the countries involved in each part of the Thematic Survey and the number of addresses of institutions each country delivered. The final date for the first three parts of the TS was the end of April 2013. Since COST B started later and the Czech Republic joined the team later, the deadline was set at the end of May. The largest number of responses was received for the ACCESS questionnaire, were 24 of the 36 institutions completed the questionnaire. The response for SIZE was also quite high: 53%. However, for all parts of the Thematic Survey the absolute number of responses was lower than the targets set in advance.

Table 3.2 shows how the absolute numbers of response are divided over the participating countries and Table 3.3 shows the same for the different types of institutions (COST A is not included anymore due to too few responses).

Table 3.2: Number and percentage of response - countries

COST B	N	%	SIZE	N	%	ACCESS	N	%
Belgium	2 (11)	18%	Austria	7 (14)	50%	Hungary	4 (5)	80%
Greece	2 (42)	5%	Estonia	6 (13)	46%	Spain	7 (9)	78%
Lithuania	5 (10)	50%	Finland	8 (13)	62%	Portugal	5 (10)	50%
Netherlands	7 (29)	24%	Germany	10 (11)	91%	Slovakia	0 (5)	0%
Poland	3 (26)	12%	Ireland	4 (12)	33%	Sweden	8 (8)	100%
Switzerland	2 (5)	40%	Luxembourg	14 (25)	56%			
Czech Republic	13 (50)	26%	Slovenia	6 (15)	40%			
UK	1 (8)	13%						
Total	**35**			**55**			**24**	

A few countries show a rather high absolute number of returned questionnaires (Czech Republic, Germany, Luxembourg, and Sweden). For some countries the percentage of returned questionnaires as compared to the number of invited institutions is high (Germany, Finland, Hungary, Spain, and Sweden).

Table 3.3: Type of institution

Type	COST B	SIZE	ACCESS
National archive	2	3	3
Other archive/Records office	5	11	4
Archives	***7 (20%)***	***14 (26%)***	***7 (29%)***
Museum of art	5	5	0
Museum of archaeology, history	0	5	1
Museum of natural history and natural science	0	0	0
Museum of science and technology	1	0	1

Type	COST B	SIZE	ACCESS
Museum of ethnography and anthropology	0	1	0
Other type of museum	5	9	2
Museums	***11 (31%)***	***20 (36%)***	***4 (17%)***
National library	3	6	5
Higher education library	2	5	2
Special or other type of library	9	10	6
Libraries	***14 (40%)***	***21 (38%)***	***13 (54%)***
Audio-visual or broadcasting institution	1	0	0
Film institute	1	0	0
Institution for Monument Care	1	0	0
Other	***3 (9%)***	***0 (0%)***	***0 (0%)***
Total	**35**	**55**	**24**

Since the numbers are small the results are not representative for the type of institution. The results only give indications and qualitative information about the questions.

Table 3.4 shows that in the Thematic Survey institutions with a relatively high annual revenue budget (between 1 million and 10 million Euros) seem over represented. This may be related to the way institutions were selected to take part. Stakeholders were aware that the survey was aimed at collecting in depth data. Therefore it would be understandable that they would tend to invite larger institutions where, on the whole, collection and information management are organised in a more systematic way. Because of the small numbers, however, we cannot determine the exact degree of bias in the results. It is most important is to keep in mind that the information that was gotten from this Thematic Survey is qualitative. There was no aim to get representative data, as in the Core Survey 1. The aim was to test particular questions in order to improve them, and to show the qualitative outcomes of some additional questions.

Table 3.4: The institution's annual revenue budget

Budget	COST B	SIZE	ACCESS	Total
< 10,000 €	1	1	2	4
10,000-50,000 €	1	2	2	5
50,000-100,000 €	1	3	2	6
100,000-500,000 €	3	9	1	13
500,000-1,000,000 €	6	5	3	14
1,000,000-10,000,000 €	17	23	8	48
>10,000,000 €	6	10	6	22
No answer	0	2	0	2
Total	**35**	**55**	**24**	**114**

4 ACCESS

The ACCESS part of the Thematic Survey was aimed at reviewing, testing, and where possible improving the questions on the topic of the *Access to and Use* of digital heritage collections in the EU. The Thematic Survey was also designed to see if some additional aggregated information on the topic could be collected with fairly straightforward questions.

The four questions in the Digital Access section of ENUMERATE Core Survey 1 were:

> *[19] Does your organisation have an explicit (written) policy regarding the use of your digital collections?*
>
> *[20] Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?*
>
> *[21] If yes, how?*
>
> *[22] Please indicate estimated percentage of all the digital objects you have that are and/or will be accessible through the access options*

Questions [19] and [20] were simple Yes/No questions, whereas questions [21] and [22] presented a selection of pre-defined answers, supplemented with an option to mention 'Other'.

The question about written policy documents showed that in a sample of 1488 institutions 31% have an explicit policy regarding the use of the digital collections. The second question in this section produced an average of 42% of institutions (n=1495) measuring in some way the frequency of accessing their digital collections. The answers to questions like these can become more useful if they are supplemented with information about the intended users, the conditions of use, and specific monitoring practices.

In the reviewing of Core Survey 1 several comments related to the ambiguity that may exist in answers about such things as the "(written) policy" and "measuring the number of times digital metadata and/or digital objects are being accessed". The general criticism concerned the high variability that is possible in the underlying reality. A written policy could be a frequently used and updated document of a hundred pages giving exact regulations and guidelines for offering access to an institution's digital collections. It could also be a one page only text, written a few years ago and 'lost in a drawer' somewhere.

Something similar could be the case with the monitoring of the number of times a digital collection is accessed. How valid is the monitoring methodology? Is it done in all of the institution's publicly available collections? What is publicly available? Is measuring organised at regular time intervals? How often then? What is being done with the results of measuring?

The general impression is that monitoring access is often only loosely related to activities aimed at improving the situation in the individual institution.

Since the ENUMERATE Core Survey targets both the collection of aggregated data on a EU scale and offers institutions insight into their own practices, the second issue of the survey could be made more actionable. But in as far as the ENUMERATE series of surveys is intended to offer insight into trends related to digital heritage collections, the questions of the first Core Survey must at least be repeated in Core Survey 2. The research here made it clear where minor adjustments are possible.

In the Thematic Survey on Access the starting point consisted of questions focussing on the use of digital collections from the ENUMERATE Core Survey (2012). These questions were augmented with those that were developed from suggestions by the Specialist Meeting on Use/Access (Madrid, 6th June 2012) in mind. Other sources are recent surveys in the field, e.g. the ITHAKA S+R Survey on Sustaining Digitized Special Collections, and especially a small survey that was conducted in the Netherlands by Henk Voorbij (2009).

An online test version of the questionnaire on USE/ACCESS, was developed by Panteia and links to the questionnaire were distributed among about institutions in 5 EU member states.

Below the results of the Thematic Survey on ACCESS are discussed along the headings in the questionnaire that were used: *Policy*; *Access to online digital collections*; *Conditions of use*; *Monitoring*; and *Evaluative remarks*.

4.1 Policy

11 of 24 institutions have an explicit written policy regarding the use of their digital heritage collections. In 7 of the 11 cases this written policy document mentions *and* elaborates upon specific types of use of the digital heritage collections.

Collections are made accessible to the public for various types of use. The table shows the types of use that are considered most important by the institutions. It is clear that educational purposes and personal cultural development are considered the most important. Commercial trading is not an important driver for the institutions to have digital collections.

Table 4.1: How important is each of the following types of use for your institution's motivation to make digital heritage collections publicly available online? (10 = very important …1 = not important at all)

Type of use	Average
Educational purposes	9.0
Personal (cultural) development	8.8
Enjoyment	6.4
Creative reuse/Remix	4.6
Religious and commemorative use	3.7
Commercial trading	2.5

4.2 Access to online digital collections

The object types which are most frequently mentioned as being made accessible online, either as digitally reproduced or born digital, are the visual 2D resources, followed by text based resources and archival resources. Of all institutions (n=24) this is mentioned by 83% (20 institutions).

Table 4.2: What object types in the digital heritage collection of your institution, both digitally reproduced or born digital objects, are accessible online?

Type	%	N
Visual (2D) resources (such as drawings, engravings/prints, maps and ground plans, paintings etc)	83%	20
Text based resources (such as books, serials and other text based resources)	79%	19
Archival resources (such as archives: government documents, government archives, other archival material)	58%	14

Digital interactive resources [exclusively digital](such as databases, digital (3D) designs, etc)	46%	11
Time based resources (such as audio files: music, speech & other, digital audio books, film	29%	7
3D Movable objects (such as 3 dimensional works of art, archaeological man-made objects, coins, etc)	17%	4
Geography based resources (such as landscapes, archaeological sites, monuments and buildings etc)	13%	3
Natural resources (such as natural inert specimens and natural living specimens)	8%	2

Specific object types that are made accessible digitally most often are: newspapers and other serials, Digital research files (incl. GIS files), Photographs, other archives and other printed books.

Table 4.3: What kind of descriptive metadata/cataloguing data are you providing online? (n=23)

	%
Metadata about items in all digital collections, including offline digital collections	89%
Metadata about items in online accessible digital collections	56%
Metadata about items that are not digitally available	11%
Other metadata	11%
No metadata at all	0%

Table 4.4: Please indicate the estimated percentage of all the digital objects you have (digitally reproduced and/or born digital) that are and/or will be accessible through mentioned access options (currently and in 2 years from now) (n=21)

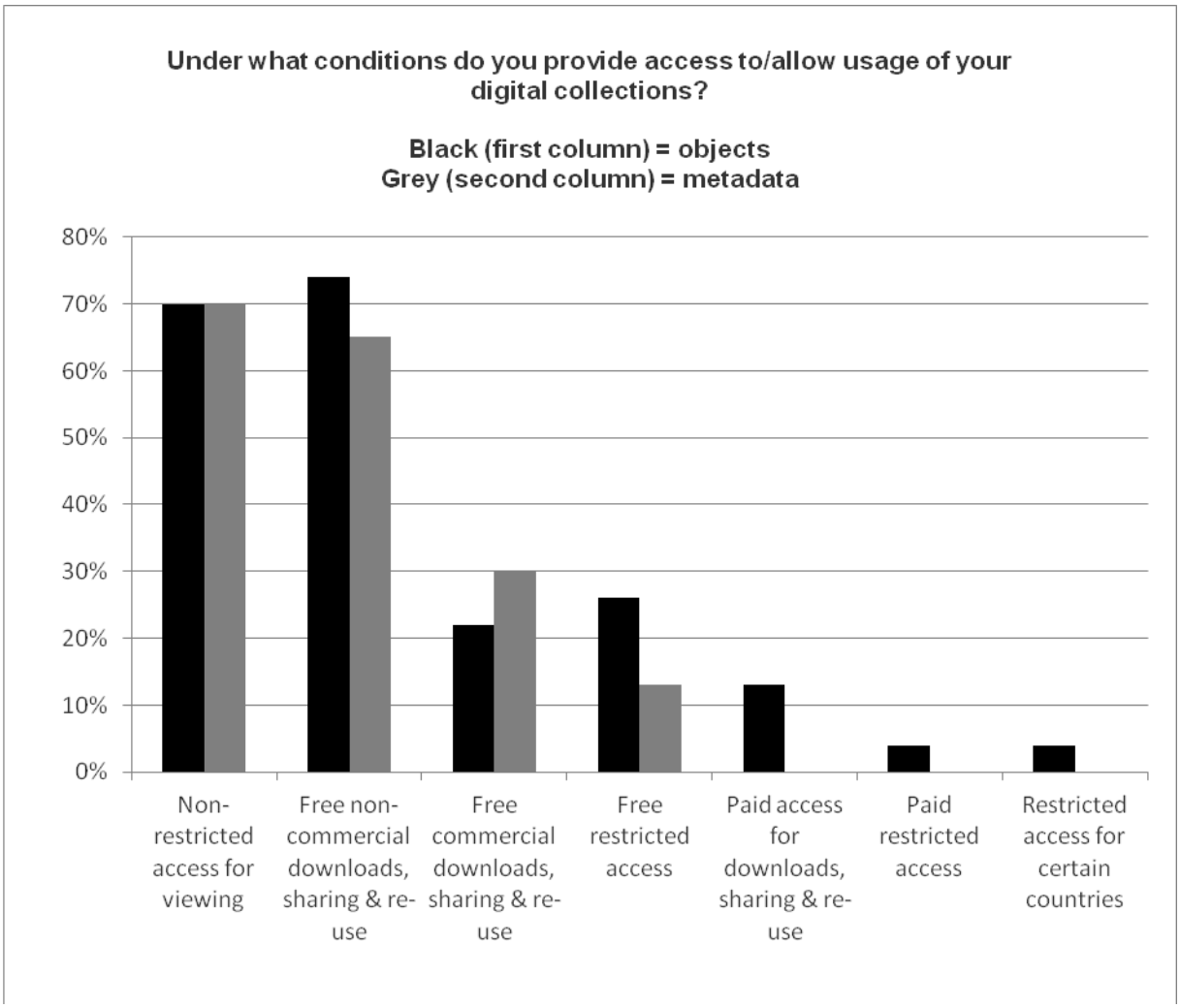
% of digital objects	**Currently Available (a)**	**Within 2 years (b)**	**Difference**	**Core Survey 1 (a)**	**Core Survey 1 (b)**	**Difference**
Institutional website	**52%**	**58%**	*6%*	**31%**	**47%**	***16%***
Offline	48%	48%	*0%*	49%	61%	*12%*
National aggregator	27%	39%	*12%*	**22%**	**37%**	***15%***
Europeana	25%	38%	*13%*	**15%**	**31%**	***16%***
Thematic aggregator	11%	13%	*2%*	14%	24%	*10%*
Institutional API	6%	11%	*5%*	12%	23%	*11%*
3rd party API	6%	7%	*1%*	6%	13%	*7%*
Social media platforms like *Flickr* or *Facebook*	6%	8%	*2%*	-	-	-
Wikipedia	3%	4%	*1%*	3%	7%	*4%*
Something else	4%	9%	*5%*	16%	28%	*12%*

The platform that is mentioned the most for the presentation of the digital collections is the own institutional website, with an expected improvement of 6% in the next two years. A large improvement is expected for Europeana (from 25% now to 38% within two years).

We added the results of the Core Survey 1 in the last three columns. The expectation in Core Survey 1 was that the highest improvement would be made in the coming two years with the institutional website, Europeana and National aggregators. The Thematic Survey shows that the most improvement is now expected from Europeana and the National aggregators. Improvement seems to have been already made by the institutional website (31% - 47% - 52%). However, these are only indications, since the composition of the sample in the Thematic Survey was smaller and there is a bias towards larger institutions.

4.3 Conditions of use

Figure 4.1: shows the answers to the question 'under what conditions do you provide access to / allow usage of your digital collections?



The percentages in Figure 4.1 refer to the % of all respondents that checked out a certain option (multiple options were possible).

In the outcomes it is interesting to see that the majority of institutions provide non-restricted access to both the metadata and the objects, which are in many cases also available for free as non-commercial downloads. No respondent charges money for getting access to their metadata. For commercial downloads, some institutions allow free downloads of their metadata, but not of their digital objects.

Evaluative remarks:

- *We allow usage of our digital collections for free non-commercial downloads, sharing and reuse except in the case of collections are subjected to author copyright.*
- *At the moment we do not have any specific regulation abut access to / allow usage of our digital collections.*
- *I suppose metadata means technical data about the digital image, such as size, filename, file format, resolution etc.*[5]
- *Downloads are only partially possible, but low resolution screenshots can be used for non-commercial purposes.*
- *There are different levels of access and restrictions for our digital collections: Free general access (for private use, for public use standard taxes apply) except for: -Restricted access+ only viewing for a specific collection of modern graphic material (only thumbnail available and full viewing only at the Library's premises); -Restricted use for audio recordings (audio streaming, not downloadable).*
- *Only copyright free objects are published for external use, all bibliographic metadata is freely accessible.*

4.4 Monitoring

18 institutions (75%) measure the number of times digital metadata and/or digital objects are being accessed by their users, 5 organisations do not. All organisations that measure use web statistics and 8 of the 18 use database statistics.

Table 4.6: Which of the following methods does your organisation use to measure the number of times digital metadata and/or digital objects are being accessed?

Method	%	N
Web statistics	100%	18
Database statistics (if not included in Web statistics)	44%	8
User studies	11%	2
Other	11%	2

Tools being used in the organisations for measuring access are:

- Google analytics (13 institutions);
- Webaudit (1);
- database and in-house developed tools (1);
- Advanced Web Statistics 7.0 (1);
- the ILS own statistical module (1);
- AWStats (1);
- We use proprietary tools (statistics) of our integrated library system (DIGIBIB) (1);
- Logs (1);

[5] Unfortunately we did not mean technical metadata in the first place. Metadata refers especially to the descriptive metadata. So presumably we did not define this in the proper manner.

- Database reports (1).

Table 4.7: How often do you analyse the results of your access/use monitoring activities?

How often	%	N
On a weekly basis	6%	1
On a monthly basis	39%	7
On a yearly basis	39%	7
Other (please specify)[6]	17%	3
Total	100%	18

The outcomes of access/use data analysis are used in the following ways:

- *To know use percentages and other performance indicators;*
- *To measure the number of our users downloads, and to know what types of records our users download;*
- *Mainly to monitor number of users and see trends in our search tool on the museum website;*
- *Official statistics;*
- *Annual report and development;*
- *For defining the ways of further developments;*
- *In relation to the access data, we develop strategies about the Digital Network;*
- *To know how to better exploit collections;*
- *To value our efforts about collections digitisation, the accessibility to the Digital Network, research has been done made by users, etc;*
- *We analyze what type of archival documents/objects are most searched/used. This helps us to make decisions for coming digitalization;*
- *Evaluating access policies;*
- *Decision-making on content, use and trends, and strategy evaluation on digitisation lines and access;*
- *Evaluation of impact; source for convincing the rights owners of other film titles to give us the necessary permission to deliver the digital objects of the materials held in our archives;*
- *Projects and activities planning;*
- *Planning necessary adjustments to the institutional repository;*
- *Internal communication (performance measures) and external;*
- *The analysis of access is used as a selection criteria;*
- *We check the outcome of the data analysis records and their visibility in the network.*

[6] The 3 'other' answers are = (1) on demand, (2) if it is necessary and (3) randomly.

4.5 KPI's (Key performance indicators)

The following KPI's are applied based on monitoring (access / use) data:

KPI's and other

- *ISO 11620;*
- *In-house KPIs;*
- *Indicators about the data our items have been look up;*
- *The increase of the Digital Network (referred to number of participating museums or number of digital heritage objects);*
- *Our KPI's are included in the General Strategic Planning of the Ministry of Culture.*

Total number of visitors

- Total sum / number of visitors / users (4);
- Total sum / number of unique visitors / users (3x);
- Total sum of visited websites and how many of these are visited from abroad;
- Total number of views;
- Duration of session, traffic (downloads).

Access

- Number of times the site was accessed;
- Number of times the digital objects were accessed;
- Access digital records available;
- Access to digital objects;
- Access to viewed pages.

How does your organisation report on these KPI's:

- *KPI's are used for internal reports but we do not make available to public;*
- *Results of our KPI's aren't published on the institutional website;*
- *Yes, in the annual report to the ministry of culture. The annual report is also published on our website;*
- *Integrated into the annual report, reported in special newsletter (for public external), balance sheet activity;*
- *Monthly on the institutional website, and quarterly + annually reporting;*
- *The indicators are published in the institutions annual report and also in the institutional website.*

4.6 Evaluative remarks of the respondents

On average it took the respondents 60 minutes to fill out the questionnaire. (minimum: 15 minutes, maximum: 3 hours). There were two respondents who said that it took them two days to fill out the questionnaire. Almost all respondents needed to look up things before filling out certain questions.

5 SIZE

There were two problematic issues in the first Core Survey related to measuring the size and growth of digital collections that we wanted to test and improve upon in the Thematic Survey (TS). The first is the complexity of collecting concrete figures. If questions are too high-level it becomes difficult to get useful results from the analysis, but if the questions are too detailed the number of respondents drops dramatically. The second was that data on the state of affairs in the collection of born digital materials was generic, and not quantifiable.

The question on the size of collections was too detailed

In Core Survey 1 the size of digital collections was roughly estimated as a percentage of the analogue collections. Two questions were posed in this respect:

> *[13] Estimate the percentage of your heritage collections that has already been digitally reproduced*
>
> [*14] Estimate the percentage of your heritage collections that still needs to be digitally reproduced*

The estimates would not normally add up to 100%, since the aim is rarely to completely digitise all analogue heritage collections. Often the most popular or most significant parts of a collection are selected for digitisation, and collection items that are not unique (e.g. books, prints, and other types that occur as duplicates) may be excluded if digitisation is done by other institutions.

Both questions were supplemented with a list of high-level object types that could be checked if they were included in the estimate. It was expected that this approach would not result in a detailed account of the state of affairs. So a more thorough approach to quantifying heritage collections was offered as well as an additional option. Only about 10% of the respondents were willing to take the challenge and fill in the optional table.

The TS was used to find out if the measuring of collection size could be more precise than was realised with the two high-level questions and yet be sufficiently feasible to raise the percentage of respondents having precise figures above the 10% that was attained in Core Survey 1.

Born digital heritage materials were excluded

The Numeric project did not include questions about born digital cultural heritage. ENUMERATE introduced this type of digital heritage into the methodology, but only on a generic level, to test the waters:

> *[15] Does your organisation collect born digital heritage?*

This question was followed by an open-ended question:

> *[16] If Yes, what types of born digital?*

The results showed that already more than half of all the institutions in Europe collect born digital heritage (52%). In the Thematic Survey, we wanted to get a better understanding of how to measure the size of these collections.

Quantifying digital collections in the Thematic Survey: three approaches

It was decided to test alternative ways of measuring the size of digital collections in order to get more and better data, and to include the measuring of the size of digital born materials in the test. The approach was to have a simple measuring tool that could be used by all respondents, although the intention to include an optional question in which institutions could specify in more detail the size of their digital collection was maintained.

At the core of the Thematic Survey SIZE questionnaire is a list of object types that represent digital heritage collections. Since the aim was to assess the size of both digitised analogue and born digital collections combined, and since a list developed for that purpose could not be traced in earlier research and surveys (see ENUMERATE Deliverable D2.2), the list had to be developed afresh. The new list is a combination of the NUMERIC and the ENUMERATE Core Survey 1 list of object types and a list of born digital object types that was developed in research on born digital heritage in The Netherlands, initiated by DEN in 2009. Accompanying questions are selected to put the results of assessing collection size and growth in context.

We chose three different approaches to quantifying digital collections. To reduce the risk of impairing the willingness of respondents to take part in the survey, it was decided to invite each participating institution for only 2 out of 3 approaches. That would be enough to compare the feasibility of the measuring techniques - do they relate to the daily practice in the institutions? - and to assess the variance that results from using different methods of counting. The institutions in this part of the TS were asked to evaluate the practicability of the different methods they applied, to help us make decisions for Core Survey 2 on the appropriateness of different methods in different sub-domains of the heritage domain.

In the TS questionnaire a control question was included to ascertain that responding institutions actually did have heritage collections:

Is curatorial care for the collections of your institution part of its mission?

For most institutions (91%, n=48) curatorial care for the collections of the institution is part of the mission of the institution. Institutions where this is not the case could be excluded from some of the analyses.

Another control question focused on whether institutions actually did have digital collections. Only one responding institution answered negatively.

In the main body of the questionnaire questions were asked about collection size using the three different approaches. To have a better understanding of the situation in the participating institutions, questions were included that made local situations intelligible. First, we will present the different measuring methods and the accompanying results separately, after that we will draw some general conclusions.

5.1 Collection registration system (CRS) approach

In this fairly straightforward approach the aim was to measure the size of digital collections based on the quantities of descriptive metadata records in an institution's Collections Registration System (Collection Management System). To put the quantification questions into context some questions about local circumstances were asked.

The first question in the CRS approach was about the CRS in use, an open question. Of all 55 institutions participating 33 institutions (62%) have mentioned the name of their CRS. Most

mentioned are: MuseumPlus, ESTER and Aleph. Of these 33 CRSs 25 (76%) are online available and the URL of the online access point is given.

For all institutions that have a CRS, the percentage of the collection that is not registered in the CRS on average is 39%.

Asking for the software that runs the CRS a large variety of software was mentioned. There is no specific software that was mentioned remarkably often.

The total number of descriptive metadata records in the CRSs varied from a few thousand to more than 10 million. The sum total for all respondents was 23 million. The total number of bytes of the digital heritage collections as recorded in the CRS of the responding institutions varied from 1 GB (a small institution) to 456 TB (a large institution).

The core of the CRS approach was an extensible table where respondents could specify the quantities of object types represented in their CRS. No efforts were taken to standardise these object types, in order to make it easier to look up data in the CRS. To check the variety of object definitions an open question was included asking for the local classification system in use:

- *What classification system or list of object types is in use to classify the objects recorded in your CRS?*

Here classification systems are mentioned such as UDC (Universal Decimal Classification) (3 institutions), DDC (2), OCM (1), General standard archival description and ISAD(G) archive description standard (1) but also a lot of institutions mention classification systems and lists of object types that have been developed by themselves:

- *Own generated lists of classification (self-defined)* (4);
- *Specially designed classification system is developed using Art & Architecture Thesaurus (AAT) etc*;
- *SIRAnet classification system developed by Slovenian regional archives*;
- *Slovenian general subject glossary (Our collections are divided into 10 regions of Slovenia)*;
- *Estonian national classification system*;
- *Internal guidelines for describing archival records*;
- *Finnish description of archival material by National Archives and own rules*;
- *For Elica: self-developed Thesaurus (gender-specific) for ALEPH (no archive documents so far) : RVM*;
- *Classification system of Art Collections*;
- *Repertoire de vedettes-Matiere (RVM)*;
- *Library of Congress Subject Headings (LCSH)*.

Other answers given were:

- *We are using the following types of objects: Text, audiovisual, animation, image, facsimile, biography*;
- *We do not use any formal thesaurus as there is nothing available for toys*;

- *Original format/type + by significant collection name*.

Given the diversity in mentioned classification systems and the relatively large number of times a self-defined system is mentioned, the conclusion can be drawn that more standardization is needed here for future measuring.

In the actual question - *Please specify the quantities of the object types that are represented in your CRS* - the table did not have a fixed number of rows. In the columns respondents could specify:

- Object type;
- Number of descriptive metadata records;
- Number of digitally reproduced or born digital objects.

The instruction suggested starting with the most numerous object types. This may have caused respondents to stop adding new rows for the less numerous object types. All in all the total number of records respondents additionally specified is substantially less than expected.

The object types mentioned most often are: photographs, books, archival objects / records, maps, audiovisual / audio recordings, manuscripts, newspapers and posters. These are all categories that are represented in the classification that ENUMERATE designed as well.

5.2 Detailed measuring of digital sub-collections approach

In this second approach the aim was to measure the size of specific digital heritage collections through a detailed quantification, starting from a pre-defined table of digitally reproduced and born digital objects. It was up to the institution to select the specific collection type. The idea behind this approach was that we could tap into intelligence about the management of digital collections that the institutions considered to be important for themselves.

Institutions (n=34) were asked to indicate which object types are represented in their chosen digital and analogue collections. Most mentioned object types, by 10 or more institutions, are:

Table 5.1: Most mentioned object types in collections (Please select all object types that are represented in your digital and analogue collections, column A)

Type	N	%
Photographs	25	74%
Posters	17	50%
Video recordings	17	50%
Drawings	17	50%
Maps and ground plans	15	44%
Other printed books	15	44%
Journals	14	41%
Engravings / Prints	14	41%
Archives : Other archival records	12	35%
Databases (containing cultural heritage metadata)	12	35%

Sheet music	11	32%
Rare printed books	11	32%

Table 5.2 shows that the percentage born digital in the entire digital collection is highest concerning databases, followed by electronic books, websites etc. In Table 5.2 only material is mentioned for which the percentage born digital is 25% or higher. The percentages concern the size of the collection, so 79% of the databases in the collection of the institutions that answered the question are born digital (i.e. not digitised from analogue sources).

Table 5.2: Estimated percentage of born digital of units in entire digital collection (n=34)

	%
Databases (containing cultural heritage metadata)	79%
Electronic books (eBooks)	75%
Archives : Government documents	50%
Audio files: Music	40%
Video recordings	34%
Photographs	31%
Newspapers	30%

Types of objects with a high estimated percentage of the entire analogue collection that is already digitised, still needs to be digitised or does not need to be digitised are shown in Table 5.3. Since some born digital material was mentioned here, such as eBooks, databases and websites, we can conclude that this question was not completely clear to all respondents. We have left these answers out of Table 5.3 for that reason.

Table 5.3: Estimated percentage of the entire analogue collection that is digitized, still needs to be digitised or does not need to be digitised (n=34)

Collection type	% Digitised	% To be	% Does not need
Other text based	50	26	24
Video recordings	46	27	27
Journals	38	12	50
Other visual (2D) resources	35	32	33
Photographs	34	38	28
Engravings / Prints	34	59	7
Rare printed books	33	50	17
Paintings	33	25	42
Drawings	32	24	44
Newspapers	30	42	29
Audio files: Music	26	53	21
Maps and ground plans	21	42	37
Film (cinematic)	20	33	48
Archives : Other archival records	13	34	53
Posters	7	59	34

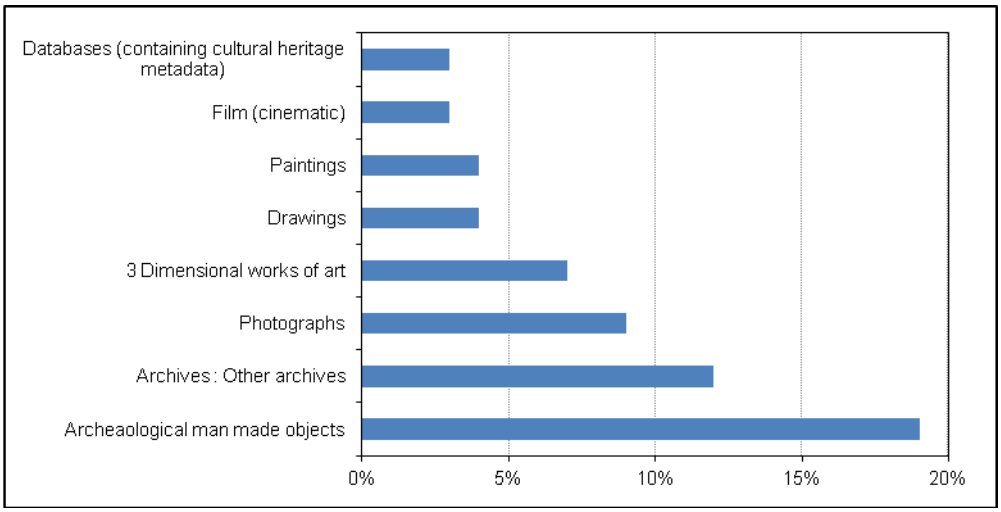
Other Manuscripts	4	86	11
Other (3D) man-made objects	3	96	2
3 Dimensional works of art	2	56	42
Archives : Other archives	0	100	0
Coins and medals	0	100	0
Medieval Manuscripts	0	93	7
Sheet music	0	55	45
Audio files: Speech & other	0	45	55
Microforms / Microfilms	0	40	60
Other printed books	0	38	62
Archives : Government archives	0	10	90
Archives : Government documents	0	8	92
Other serials	0	0	100
Digital audio books	0	0	100
Other time based resources	0	0	100

5.3 Improved ENUMERATE Core Survey estimation approach

The third approach chosen here is based on the approach from the ENUMERATE Core Survey 1. It is about estimates of the size of the physical heritage collections, the proportion of it that has already been digitised and the number of objects that still need to be digitised.

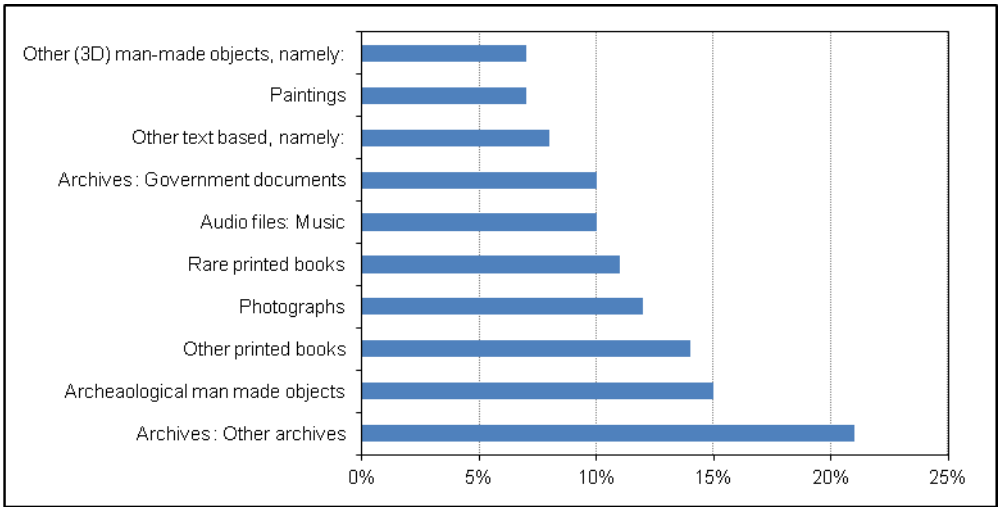
On average the institutions mention that 23% of their analogue heritage collections have already been digitally reproduced (In ENUMERATE Core Survey 1 it was 20%). The objects types mentioned most often as making up for this 23% are the object types mentioned in Figure 5.4 (here we only mention the object types that are mentioned most).

Figure 5.4: Estimate how this percentage of digitally reproduced analogue objects is distributed in percentages amongst the different object types (n=25)



On average the institutions mention that 55% of their analogue heritage collections still needs to be digitally reproduced (in ENUMERATE Core Survey 1 it was 57%). This holds mainly for the following object types (see Figure 5.5). Also here the most mentioned object types are presented.

Figure 5.5: Estimate how this percentage of analogue objects that still needs to be digitally reproduced is distributed in percentages amongst the different object types of your collections.



Born digital heritage materials are collected by 67% of institutions (23 of total n (=34); in ENUMERATE Core Survey 1, it was 52%). One third of the institutions do not collect born digital heritage materials. Most often collected born digital materials are photographs, video recordings and audio files (see Table 5.6). These are the same three main categories as in Core Survey 1.

Table 5.6: If born digital heritage material is collected, what types of born digital material does this concern?

Born digital collection (n=23)	N	%
Photographs	15	65%
Video recordings	12	52%
Audio files: Speech & other	10	44%
Archives : Other archival records	5	22%
Electronic books (eBooks)	5	22%
Film	5	22%
Audio files: Music	5	22%
Databases (containing cultural heritage metadata)	4	17%
Other digital born text documents	4	17%
Websites (incl. blogs, tweets, widgets, wikis)	4	17%
Archives : Government documents	3	13%
e-Newspapers	3	13%
e-Journals	3	13%
Digital research files (including GIS files)	3	13%
Digital (3D) designs or reconstructions of objects and buildings	2	9%
Digital maps and ground plans	2	9%

Newspapers	2	9%
Other e-Serials	2	9%
Games	2	9%
Software (customised)	1	4%
Digital art objects (including Internet art)	1	4%

5.4 Evaluative remarks of the respondents

On average the third approach proved to be the most time consuming, with an average of 80 minutes for filling out the questionnaire. The least time consuming approach is the CRS approach with on average half an hour for filling out the questionnaire.

Table 5.7A: How much time was needed to fill out the questions for the different approaches?

Approach	Number of Minutes Mean	Percentile 25	Percentile 75
CRS approach	31	10	45
Detailed sub-collection approach	52	15	60
Improved ENUMERATE approach	80	15	60

Table 5.7B: How much time was needed to fill out the questions for the different approaches related to the annual revenue budget?

	What is your institution's annual revenue budget?							
	< 10,000 Minutes Mean	10,000-50,000 Minutes Mean	50,000-100,000 Minutes Mean	100,000-500,000 Minutes Mean	500,000-1,000,000 Minutes Mean	1,000,000-10,000,000 Minutes Mean	>10,000,000 Minutes Mean	**Total Minutes Mean**
CRS approach	.	5	45	16	19	43	15	**31**
Detailed sub-collection approach	.	60	.	27	18	55	79	**52**
Improved ENUMERATE approach	.	.	22	38	23	131	61	**80**

Table 5.7B shows to which degree the size of the institution (based on the annual revenue budget) influences the evaluation of the three approaches. In general it can be concluded again that the CRS approach is the less time consuming for all sizes of institutions. However, the table shows some more details and nuances. The results show that, whatever approach will be chosen for ENUMERATE Core Survey 2 we should indicate explicitly that just this one topic of the questionnaire will take on average about 1 hour to answer.

Table 5.8: A How difficult was it to collect the necessary information to answer the questions in each of the two approaches

Approach	How difficult (Mean)	N
CRS approach	3	31
Detailed sub-collection approach	2	33
Improved ENUMERATE approach	2	34

Each respondent was asked how difficult it was to collect the necessary information to answer the questions in each of the approaches. This question could be answered on a scale form 1 = 'very difficult' to 5 = 'very easy'. The mean in Table 5.7 shows that in the CRS method on average is easier for respondents to collect the necessary information than in the two other approaches. Thus, the CRS method is the easiest and takes the least time to fill out.

Table 5.8B: Which one of the approaches would you prefer to fill out in the future issues of the ENUMERATE Core Survey?

Approach	Preferred (N)	Most faithful (N)	Preferred (%)	Most faithful (%)	Total N
CRS approach	25	22	81%	71%	31
Detailed sub-collection approach	10	12	30%	36%	33
Improved ENUMERATE approach	13	13	38%	38%	34

Also Table 5.8B shows that the CRS approach is the most preferred and it also gives the most accurate data. For reasons why each of the approaches is preferred and or most accurate, see Appendix 3.

6 COST

In the Cost survey the aim was to get a better understanding of the cost management of digital collections. Specialists in selected institutions were asked to consider the costs of acquiring and valorising one specific digital collection.

In the 2012 Core Survey one of the sections was about 'Digitisation Expenditure'. Essentially three kinds of information were collected: the total annual expenditures on *digitisation*; the number of staff, paid and volunteers, involved; and the sources of funding. In the Thematic Survey (TS) the aim was to widen the scope of the questions from only 'digitisation' to the total costs of creating or acquiring, maintaining and preserving digital heritage collections. To get more specific information per cost item the TS experimented with questions in which respondents could quantify expenditures per activity. It was clear from the beginning that the challenge would be in posing questions that would relate as much as possible to the financial systems as used by heritage institutions.

Costs were divided into:

- **Incidental costs** – defined as the costs having to do with the initial creation or acquisition of a digital collection;
- **Structural costs** – the annual costs needed for the ongoing maintenance, enhancement and preservation of a digital collection.

The survey Cost B consisted of 14 questions. Since the questionnaire aimed to improve the methodology of monitoring cost, some questions had an accompanying field for evaluative remarks. Respondents were encouraged to help with any comments that might be useful for the development of a measuring methodology.

After some general questions about the organisation and its financial management related to digital collections, the questionnaire asked respondents to focus on a specific collection and to detail the incidental and structural costs involved.

Since respondents were relatively free in choosing the specific collection, the ENUMERATE team could not affect the object types that were represented in the sample. After 35 organisations had filled in the questionnaire high level object types were included according to the following distribution:

Table 6.1: What types of digital objects are in the selected collection?

High level denominators for digital object types	No. of occurrences in the sample
VISUAL (2D) RESOURCES (such as drawings, engravings/prints, maps and ground plans, paintings, photographs)	20
TEXT BASED RESOURCES (such as books, serials and other text based resources)	13
ARCHIVAL RESOURCES (such as: government documents, government archives, other archival materials)	7
3D MOVABLE OBJECTS (such as 3 dimensional works of art, archaeological man-made objects, coins and medals)	5
TIME BASED RESOURCES (such as audio files: music, speech & other, digital audio books, film and video recordings)	5
GEOGRAPHY BASED RESOURCES (such as landscapes, archaeological sites, monuments and buildings)	2

DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL] (such as databases, digital (3d) designs, digital art objects)	2
NATURAL RESOURCES (such as natural inert specimens and natural living specimens)	0

6.1 Costs in the annual reports of memory organisations

To get an impression of the cost management at the selected institution, the Thematic Survey questionnaire contained a question about the annual reports of the heritage institutions and whether their digital collections were mentioned in particular:

> *Does the most recent annual report of your institution explicitly mention costs related to your institution's digital collections? [Yes / No / Do not know]*

Only 10 (28.6%) out of 35 respondents (n=35) answered Yes to the question. This may be taken as an indication that it may be difficult to separate the costs of digital collections from other reportable costs.

As more than 70% institutions answered negatively to the previous question the results of the follow-up question can only be very broadly indicative:

> *If you answered Yes in the previous question what does this most recent annual report of your institution mention?*

A number of predefined answering categories that could be checked were offered:

- *Growth of digital collections;*
- *FTEs involved in creating and managing digital collections;*
- *Budget related to creating and managing digital collections;*
- *Growth of digital storage (in Terabyte);*
- *Other (please specify below).*

The ENUMERATE team decided not to use an extensive listing here, in the hope that respondents would provide insight into the way the costs of digital collections are framed relative to the annual revenue budgets of memory organisations.

The set-up did not work in practice. Only two lump sum costs were mentioned: "*participation in [the] Google Art Project*" and "*overhead, exhibitions, screenings & events, media library, distribution*".

Among the predefined categories the item "*Growth of digital collections*" was never checked out. For the other options the results are in Table 6.2.

Table 6.2: Cost as mentioned in annual reports (n=6)

Cost	%
Budget related to creating and managing digital collections	83
Growth of digital storage (for example in number of Terabyte)	50
FTEs involved in creating and managing digital collections	33

6.2 The cost of specific digital heritage collections

As mentioned respondents were asked to specify the actual costs of one specific digital collection, separating incidental costs from structural costs. To make the task of breaking down costs more concrete 9 high-level incidental costs were predefined and 8 high-level structural costs. The costs were all activity based, which was clearly indicated in the questionnaire instructions[7]. The option to enter *Other costs* (*please specify: …*) was intended to learn more about the way institutions have access to financial data. As we will see the additional costs mentioned were only in part activity-based. On the whole respondents gave more detailed answers in the section on Incidental costs.

6.2.1 Incidental costs

Of the 35 institutions that participated in the Cost B survey, 33 provided specified information about their incidental costs.

The following question was posed to collect information about the incidental costs:

[13] Activity based costing: determining Incidental costs

We now invite you to estimate how much your institution has spent (in Euros) for *the first time creation or acquisition* (incidental costs) of the selected digital collection (including the cost of staff time) for each of the following activities, wherever these occurred in the institution or have been outsourced.

Leave blank the items that you cannot estimate. Enter 0 if a cost item is not applicable. If you miss any items in the table, please help us and mention these under *Other costs*.

Because *n* is low and the collections selected for this question were rather hybrid or multiform - the average number of object types per collection was 3; the median was 2; the highest number of object types was 16 - we only calculated percentages.

Because of the hybrid nature of the collections selected here - which of course reflects the practice in heritage institutions - it is not possible to relate the specified costs to specific object types.

The table below is sorted along the percentage of the specific cost as compared to *all* costs specified. The subtotal does not add up to 100% because about 20% of the costs mentioned were specified under the denominator 'Other'. (See further below.)

Table 6.3: How much has your institution spent on incidental costs of the selected digital collection, whereas these occurred in the institution or have been outsourced

COST	N	% OF ALL COSTS	% IN HOUSE	% OUTSOURCED
Analogue-Digital conversion	33	44	74	26
Web design and/or software development	17	9	21	79
Metadata creation and/or enhancement	20	9	89	11
Acquisition of digital born material	10	8	97	3
Project management	25	6	95	5

[7] This activity based costing method was a recommendation from the specialist meeting in UK (London) on Cost of digital heritage (March 2012).

COST	N	% OF ALL COSTS	% IN HOUSE	% OUTSOURCED
Selection of material for digitisation	12	2	97	3
Copyright clearance	7	1	96	4
Logistics (shipment of collection for digitisation, etc)	11	0	68	32
Collection of user generated content (UGC)	2	0	50	50
Total / mean		79	74	26

In the instructions respondents were asked to enter 0 if a cost item was not applicable and to leave the cell blank if an estimate could not be made. Based on an inspection of the survey data there is some doubt about whether all participants have followed this rule in a consistent manner. That being said, for the in-house activities, we can deduce a list of *incidental cost items* sorted from the most estimable to the least estimable:

Table 6.4: Assessment of the difficulty to estimate the costs of specific incidental activities. The lower in the list, the more difficult it is to estimate the cost item

Cost
1. Project management
2. Acquisition of digital born material
3. Analogue-Digital conversion
4. Copyright clearance
5. Metadata creation and/or enhancement
6. Web design and/or software development
7. Logistics (shipment of collection for digitisation, etc)
8. Collection of user generated content (UGC)
9. Selection of material for digitisation

The pre-defined costs were presented in the questionnaire to make the exercise in specifying costs more concrete. The simple workflow model for calculating costs here is an adapted version based on earlier and present experience with this kind of research.[8]

The option of specifying other costs than those that were pre-defined resulted in a motley collection of costs. In the small sample of this Thematic Survey there was almost no overlap in mentioned costs. Two institutions only filled in “Total costs”, without further specification. In the table below costs are sorted from the highest % to the lowest % of the sum of all costs.

[8] ‘Rekenmodel digitaliseringskosten (Rekenmodel)’, technical report, The DEN Foundation (The Hague, 2010). Refer to: **http://www.den.nl/standaard/202/Rekenmodel-digitaliseringskosten** (in Dutch) (link accessed 01/10/2013).

Table 6.5: Classification of other costs

COST	N	% OF ALL COSTS	% IN HOUSE	% OUTSOURCED
Archival system	1	4.40	100	0
Licensing (system)	1	2.93	100	0
Storage	1	2.20	100	0
Processing (staff)	1	2.20	100	0
English translations info about objects	1	1.66	0	100
Purchase photographic equipment	1	1.11	0	100
Total costs	4	0.94	46	54
Describing the collection	1	0.94	100	0
Scanning and microfilming	1	0.86	0	100
Access	1	0.73	100	0
Administration	1	0.73	100	0
Restoration costs	1	0.44	0	100
Purchase a license database program	1	0.34	0	100
Purchase of hardware and computer	1	0.32	0	100
Costs for contributing to [a web portal]	2	0.24	6	94
Other costs	1	0.23	50	50
Microfilming	1	0.07	0	100
Purchase of equipment for digitisation	1	0.07	0	100
Digital office equipment	1	0.05	100	0
Promotion/publicity	2	0.04	33	67
Auditing	1	0.04	0	100
Total / mean		**21**	**72**	**28**

It is clear that these other costs are on average either in house costs or costs of outsourcing. The overall percentages (In house 72%; outsourced 28%) are close to the percentages in the table of predefined costs (In house 74%, outsourced 26%).

Furthermore, it is noteworthy that these 'fill in the blank'-costs are only in some cases ‘activity-based’. This is true for e.g. Processing (staff), English translations [of] information about objects, Describing the collection, Scanning and microfilming, Administration, Restoration costs, and so on. Some of the highest costs are capital or purchase costs (Archival system, Licensing system, Storage).

Some of the denominators indicate cost items that were felt to be missing in the predefined list: Storage, for instance, was not on the list, but it is evident that the first time creation or acquisition of a digital collection cannot do without this activity.

A final remark must be made about the terminology. The option to enter free text descriptions of costs leads to overlapping concepts by nature and broader and narrower terms. Some of the free text denominators seem to cover items equal to or similar to the predefined denominators:

'Describing the collection' can be considered the same as 'Metadata creation and/or enhancement' and 'Microfilming' is a subclass of analogue-digital conversion.

6.2.1.1 Evaluative remarks of the respondents about specifying Incidental costs

The two questions on determining incidental and structural costs had a text field for evaluative remarks. Below is a selection of the responses given in response to the topic of specifying *Incidental costs*.

[13] Evaluative remarks to this question: (Please indicate if the above mentioned activities represent parts of your workflow. If Yes, were you able to determine faithfully the costs of the various activities? If No, were you able to break down your workflow in alternative activities and accessory costs?)

Incidental costs - remarks
The […] *project was a huge project. It dealt with 1633 vintage prints in big formats* […]. *The photo's first were cleaned and if necessary restored. They are digitized […] and then put on cardboard and preserved in Melinex and ten photographs at a time were put in acid free boxes. So it is difficult to separate between preliminary but necessary cost and the pure cost of digitizing.*
Yes. The above activities are an estimate, representing about 30% of the activity of the service.
The above mentioned activities represent parts of our workflow. We were able to determine the costs faithfully.
The above mentioned activities have become parts of our workflow after implementing of the digitization project.
Yes [above mentioned activities represent parts of our workflow] - *Yes* [we were able to determine faithfully the costs of the various activities]
Activities mentioned above do represent steps of digitisation workflow implemented in National Library. Indicated costs are based on estimations.
The costs for the internal activities and for the work carried out by companies can be defined accurately. . The time spent on the project (management, preparing material) by the partners (cantonal library) and related costs are estimates.
The activities were not part of our standard workflow but of a subsidised conservation and digitazation project […]. *The cost of metadata creation and/or enhancement was not part of this project and is ongoing We expect to finish in 2014.*
Yes, we are able to determine faithfully the costs of the various activities, except the cost of author rights (digitised is only […]*).*
The activities related to the project represent 5-10% of workflow activity in 7 persons. Related costs are thus estimated at 800 Euros per month.
The above mentioned activities represent parts of our workflow.

6.2.2 Structural costs

In the second part of the specification of activity based costs respondents were invited to consider the *structural* costs of the selected digital collection. The number of respondents here was considerably lower than in the first part of the cost specification section: N=19. Furthermore some of the answers seem to be incomplete. Below we will briefly reflect on possible explanations. But first the data we collected will be summarised.

This is the question posed:

[14] Activity based costing: determining Structural costs

For the past fiscal year, please estimate the structural costs for the selected digital collection. Fill in the form for each cost item that can be estimated, wherever it occurs in the institution or is outsourced.

Leave blank the items that you cannot estimate. Enter 0 if a cost item is not applicable. If you miss any items in the table, please help us and mention these under Other costs.

Again only calculated percentages are shown in the table below and for the same reason as mentioned above no absolute costs per object type can be calculated.

The table is sorted along the percentage of the specific cost as compared to *all* costs specified. Here the costs mentioned under the option 'Other' are not included.

The subtotal does not add up to 100% because about 8% of the costs mentioned were specified under the denominator 'Other'. Because for Other costs there was no option to separate In-house from Outsourced cost the sum total for these does add up to 100%.

Table 6.6: How much has your institution spent on structural costs of the selected digital collection, whereas these occurred in the institution or have been outsourced.

COST	N	% OF ALL COSTS	% IN HOUSE	% OUTSOURCED
Archiving (storage, including backups)	18	44	94	6
Licensing	3	21	100	0
Project management	8	7	99	1
Costs related to giving access to the digital collection	13	6	90	10
Activities concerning the (long-term) preservation of the digital collection	6	6	99	1
Editorial (including content selection and updating)	6	5	100	0
User outreach and support	7	4	97	3
Usage analysis (including surveys, interviews, and other activities	4	0	100	0
Total / Mean		**93**	**96**	**4**

As in the section on incidental costs respondents were asked to make a difference between not applicable cost items and cost items for which an estimate could not be made. Here is the listing of the structural in-house activities, again sorted from the most estimable to the least estimable:

Table 6.7: Assessment of the difficulty to estimate the costs of specific activities. The lower in the list, the more difficult it was to estimate the cost item.

Cost
1. Archiving (storage, including backups)
2. Licensing
3. Costs related to giving access to the digital collection
4. Project management
5. User outreach and support (I)
6. Editorial (including content selection and updating)
7. Usage analysis (including surveys, interviews, and other activities)
8. Activities concerning the (long-term) preservation of the digital collection.

Of course we should be cautious with the outcomes presented here. Both the number of respondents **and** the number of cost item specifications was substantially lower in the section on structural costs. It would be tempting to relate this to the transparency of the accounting system, but it could just as well be caused by some sort of *survey fatigue*. Unfortunately in the time frame of the Thematic Survey we are not able to unravel this any further.

Evaluative remarks of the respondents about specifying *structural costs*

Although the number of respondents was substantially lower in the second part of the survey on costs the evaluative remarks were on the whole surprisingly similar. Some institutions are able to specify the structural costs of a specific collection. For other institutions specifying such costs was an impossible assignment. A positive sign may be that the cost categories presented were considered by several respondents to represent the workflow in their institutions. Below is a selection of the responses given in response to the topic of specifying *structural costs*.

[14] Evaluative remarks to this question: (Please indicate if the above mentioned activities represent parts of your workflow. If Yes, were you able to actually determine the costs of the various activities? If No, were you able to break down your workflow in alternative activities and accessory costs?)

Structural costs - remarks
All activities represent parts of our workflow.
The [digital collection selected here] *is a small database. We are busy now with a project with* [a partner organisation] *to try to put the* [mentioned digital collection] *in the* [Cloud]*. Maintenance costs (now paid to our own IT department) will drop then to may be 1000 EURO per year.*
Yes. The above activities are an estimate, representing about 30% of the activity of the service.
The above mentioned activities represent parts of our workflow. We were able to actually

determine the costs of the various activities.
Yes, [we were] *able to determine costs of activities.*
We tried to fill in the survey on the cost of digital collections, but we simply don't have all needful information about the costs, we don't do such specific costing, that is why we are not a good respondent to this survey and couldn't fill in it.
It is not possible to estimate the structural costs for the past fiscal year as the work flow was not even during 3 years. So the costs given above are for 36 months. The above mentioned activities have become parts of our workflow after implementing of the digitization project.
Yes [these costs are part of our workflow] - *No* [we were not able to actually determine the costs of the various activities].
Above mentioned collection was digitized by a special project, the costs were planned in the budget of the project. The data storage and web providing (potential other costs) were arranged within other activities by our institutions IT department and the costs were not calculated.
There have been costs, but at this point it is not clear, we cannot deliver these costs.
Yes these activities represent our workflow but it was difficult to estimates the costs.
All costs are in-house but no amounts are given.
Usage analysis and editorial work were not carried out last year but will be started in 2013.
To an extent the activities are part of our workflow. For some activities it proved impossible to attribute exact costs to this specific archive/project.
We cannot estimate these costs; they were not followed during the past fiscal year.
Most of the work (selecting/preparing material, adding metadata) was done by volunteers.
Above mentioned activities represent parts of our workflow.
[…] *I would like to explain the lack of numbers, I cannot specify how much time and money goes into the management of this particular collection, since it is only part of what we do.*

6.3 Conclusions

- The exercise to specify the costs of workflow activities ('activity based costing') made clear that some of the institutions were able to fill in concrete quantities. For others the exercise was too difficult or did not relate to their daily practice and local circumstances.
- The gross percentages calculated for the estimates of costs give an impression of how costs are distributed among the activities related to digital collections, but firm statements cannot be deduced.
- The test in separating In-house costs from outsourced costs gave some insight in the distribution of these high-level costs. For instance, the structural costs are almost entirely spent on in-house activities.
- In all, due to the limited response and the lack of a harmonised terminology, we will have to be cautious when interpreting the results presented here.

7 Improvements for Core Survey 2 and overall conclusions

There were two aims for the Thematic Survey:

- Testing certain questions in the Core Survey 1 so we can change, or add, them in Core Survey 2;
- Showing qualitative outcomes of additional questions (not included in the Core Survey 1).

As the number of respondents was lower than anticipated, some figures collected in the Thematic Survey, can only be regarded as indicative. Comparisons with the results from Core Survey 1 show some striking similarities, but also some differences that cannot be explained based on the data collected.

Relevant outcomes for the ACCESS questionnaire

Some interesting qualitative outcomes of additional questions that did come out of the Thematic Survey are:

- The question on 'type of use' (see Table 4.1);
- The question 'How often do you analyse the results of your access monitoring activities' (see Table 4.7).

Possible changes for the Core Survey 2 based on the outcomes of the ACCESS questionnaire are:

- For the question '% of digital objects accessible' (see Table 4.4) a new answering category can be added in Core Survey 2, namely: 'social media platforms like Flickr or Facebook'.
- For the question 'methods of use to measure the number of times digital metadata and/or digital objects are being accessed' (see Table 4.6) an additional question can be added in the Core Survey 2: When using web statistics, which are used? (Google Analytics or other?)

Relevant outcomes for the SIZE questionnaire

Some interesting outcomes of this part of the survey:

- Even though the sample is quite small, the dataset gathered seems to confirm most of the outcomes from the ENUMERATE Core Survey 1, both about the size of the digital collections and the popularity of specific collections types for digitisation.
- We have also made some progress in better understanding the composition of born digital heritage collections across Europe.

Possible changes for the Core Survey 2 based on the outcomes of the ACCESS questionnaire are:

- The survey proved that the CRS approach was the most time effective for the respondents. However, the output from this approach turned out to be not as 'rich' as in the other two approaches. This needs further discussion in the ENUMERATE team.

- There was little discussion about the classification that ENUMERATE created to designate digital heritage collection types, so this improved classification can be used in future surveys.
- Use 'descriptive metadata' instead of 'metadata'.

Relevant outcomes for the COST questionnaire

Some interesting outcomes of this part of the survey:

- A minority of institutions report separately on their digital activities in their annual reports. If this does not change, it seems unlikely that future surveys will be able to collect really big datasets on costs of digitisation.
- Although the sample is small, there seems to be a trend that the cultural heritage institutions in Europe tend to spend more incidental costs on outsourcing activities than structural costs. Long term management of digital collections is very much an in-house activity.
- There is no one way that institutions record their costs. Some institutions use activity based costing, while others use another method.
- Incidental costs could be identified, while structural costs were much more difficult to identify.
- Institutions use a large variety of different categories of costs. Some new categories mentioned are activity based while others are not activity based.
- We understand better now which cost types can be more easily assessed by some institutions than others.

Possible changes for the Core Survey 2:

- Based on the outcomes of the COST part in the Thematic Survey we will make adjustments to the question on expenditure. In the Thematic Survey we had defined a list of both structural and incidental cost types. Even though some institutions seemed to have some problems with this distinction, due to lack of a harmonised vocabulary, we will continue to use these cost types in the Core Survey 2 and ask for expenditures. The next question will be on making a division of the total expenditures in structural and incidental costs. This can be an estimate in percentages.

Overall conclusions

As the Thematic Survey was intended to test and improve the methodology with a smaller sample of institutions, there are few general conclusions that can be drawn about the status of digital cultural heritage in Europe. Most findings will feed into the preparations of the 2nd ENUMERATE Core Survey. The most important conclusions for this are:

- Although the questionnaires were rather complex and time consuming they give us useful input for Corey Survey 2 and beyond. On average they took a lot of time to fill out, even from institutions that were very motivated to participate in the Thematic Survey. There were some that did not fill out the questionnaire in the end although they had said they would do so. So careful consideration is needed to find the balance between on the one hand a questionnaire that can provide rich data to better understand the progress of digital cultural heritage in Europe, and on the other hand an approach that does not

scare away institutions from participating because of complexity or because participation is time-consuming.

- It is clear that these other costs are on average either in-house costs or costs of outsourcing. The overall percentages (In house 72%; outsourced 28%) are close to the percentages in the table of pre-defined costs (In house 74%. outsourced 26%).

Overall: the Thematic Survey gave us insight into what we should not ask, what we can ask in a different way and what we can change to the Core Survey 2. Also the outcomes give input to further discussions within the ENUMERATE team in the area of measuring the status of digital cultural heritage in Europe.

Appendix 1 Report meeting Digital Preservation (17-01-2013)

Monitoring and Measuring Digital Preservation practices in the EU cultural heritage domain

On the topic of digital preservation, it was decided not to set up a questionnaire as part the Thematic Survey, but to organise a focus group with some major stakeholders of digital preservation of digital heritage in Europe. This decision was based on the common knowledge that digital preservation practices are not yet widely spread in the cultural heritage domain. The ENUMERATE team also felt the need to discuss with some of the frontrunners what would be a realistic scope for measuring the progress of digital preservation and also to see if and how activities could be aligned.

This focus group meeting on measuring the progress of Digital Preservation took place on 17 January 2013 in The Hague, with 11 representatives from National Libraries, National Archives, Prestocentre, EU-projects on Digital Preservation, JISC, PACKED, DEN and ENUMERATE. A full report of this meeting is available on the website of ENUMERATE (www.enumerate.eu). This paragraph presents the most important outcomes of the meeting.

The experts agreed with the ENUMERATE Team that for many institutions a questionnaire on digital preservation would come too early. Digital Preservation is still very much an area of research, the level of implementation is not yet very high.

It was agreed that it would be useful to (collectively) build up a practice to collect metrics about Digital Preservation. There are three main incentives to do so:

1 For internal use in cultural heritage institutions: internal accountability; self assessment; measuring the progress that the institution made compared with its strategies; making the case internally to raise awareness and involvement;

2 For comparison with similar institutions (e.g. comparing or maybe even benchmarking cost, investment, workflows, etc.);

3 For public accountability: metrics can help to make the case for digital preservation with governments, the EU, even the general public. Contribute to Evidence based policies.

However, before measuring can be put into practice, several issues need to be addressed first. There is for instance not a clear-cut terminology accepted throughout the cultural heritage community, or even within the same institution for that matter: "Ask 3 persons in one institution for the costs of digital preservation and you get three different answers depending on different definitions." Also, on a conceptual level, it is not clear yet, what the metrics should express. Do we want all cultural heritage institutions to have their own digital preservation strategy? Or should it suffice for smaller institutions to rely on the strategies of frontrunners in this area? And there are also differences across domains and sectors: is it really useful to aggregate and compare metrics about digital preservation practices at audio-visual archives with those at libraries? Questions like these need to be answered first before questionnaires can be sent out to the entire cultural heritage domain.

The Focus Group discussed some possible questions that could be incorporated in future questionnaires. It was agreed that the questions needed to be not too abstract or high-level, but close to the institutional practices. Key area's that should be addressed are:

- The choice for open or commercial solutions;
- The willingness to collaborate;
- Self assessment of the 'digital preservation maturity' of the organisation;
- Workflows and tools.

Finally, some other monitoring activities of Digital Preservation were identified: PrestoCentre (incl. Presto4U project), PARSE.Insight, APARSEN Survey, ERPANET survey, 4C, DigCurv, NESTOR, PLANETS and the work done by Neil Beagrie. It was agreed that it would be better if ENUMERATE would tap into that research instead of starting anew. An analysis of available research would already be a big step towards understanding the bigger picture of digital preservation in Europe.

The full report of the meeting is available at:

http://www.enumerate.eu/fileadmin/ENUMERATE/meetings/ENUMERATE-FocusGroup-on-Digital-Preservation-Monitoring_TheHague_20130117_v01.pdf

Appendix 2 Questionnaires: Access, Size and Cost

ACCESS / USE

In this small scale survey the aim is to get a grasp of the access and use of digital collections. In more detail, the questionnaire we developed focuses on:

- Online access: which part of your collection is available online and what are the platforms where content is presented to the public?
- The development of online user services: has your institution developed any websites, apps, etc.?
- The measurement of online user activities: what methods and tools does your institution use for collecting usage metrics?
- Licensing and legal issues regarding the online collection: how much does your institution spend on clearing copyright?

The survey consists of 22 questions. Since the questionnaire aims to improve the methodology of monitoring the access and use of digital heritage collections, some questions have an accompanying field for evaluative remarks. Please help us with any comments that may be useful.

The time needed to fill in the questionnaire will depend highly on the availability of management information about your digital collections. If this information is in place, we estimate that answering the questions will take about 30 minutes of your time.

We hope you will be able to complete the questionnaire by March 15th. The questionnaire is designed in such a way that you can pause and return at any time to continue.

If you have any further questions, please contact your national coordinator, or:

Natasha Stroeker: enumerate@panteia.nl

Best Regards, the ENUMERATE Team

SECTION 1/6 - ORGANISATIONAL INFORMATION

[1] Name of institution / organisation *:

(Please provide the full, legal name of the company or institution about which you are answering the questions in this survey.)

[2] Type / Domain of institution / organisation *

Please specify the primary heading under which your company or institution classifies itself:

- National archive
- Other archive/Records office
- Audio-visual or broadcasting institution
- Film institute
- Museum of art
- Museum of archaeology, history
- Museum of natural history and natural science
- Museum of science and technology
- Museum of ethnography and anthropology
- Other type of museum
- National library
- Higher education library
- Special or other type of library
- Institution for Monument Care

[3] Your name:

(The full name and title of the person completing this survey.)

[4] Your telephone number and/or Skype contact details:

(The primary phone number (e.g. +44 0123456789) and/or the Skype details of the person completing this survey.)

[5] Your role in the institution:

(The position (job title) of the person completing this survey.)

[6] Your e-mail address:

(The primary e-mail address of the person completing this survey.)

[7] What is your institution's annual revenue budget?

Please choose only one of the following:

- < 10,000 €
- 10,000-50,000 €
- 50,000-100,000 €
- 100,000-500,000 €
- 500,000-1,000,000 €
- 1,000,000-10,000,000 €
- >10,000,000 €

(Provide the annual budget for the *entire* institution, for instance as indicated in the last set of published accounts. The budget may include government funding, project funding, revenues from the organisation's primary or commercial activities, etc. If choosing between two of the options is difficult (like when the budget is exactly 50,000 €), please choose the option with the lower estimate.)

SECTION 2/6 - POLICY

[8] Does your organisation have an explicit (written) policy regarding the use of your digital heritage collections?

- Yes
- No
- Do not know

(Answer this question with "yes" if your organisation has a formal document detailing which digital materials are accessible to whom and what the terms and conditions of this accessibility are.)

[9] If it is available online, what is the URL for this policy document?

(Please specify the URL for the above mentioned policy document if it is publicly available online.)

O not available online

[10] If you answered Yes to question #8, does this written policy document mention *and* elaborate upon specific types of use of your digital heritage collections? (Refer to question #11 for an overview of examples of types of use.) (Specific types of use could be: Creative re-use/Remix, Personal (cultural) development, Educational use, etc.)

- Yes
- No
- Do not know

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[11] Collections are made accessible to the public for various types of use. How important is each of the following types of use for your institution's motivation to make digital heritage collections publicly available online?

types of use	1	2	3	4	5	6	7	8	9	10
Creative reuse/Remix										
Commercial trading										
Personal (cultural) development										
Educational purposes										
Enjoyment										
Religious and commemorative use										
Other types of use (please specify)										

(Using a 10-points scale - where 1 equals "not at all important" to 10 "very important" - please select only one number per row.)

Other types of use are:

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS

[12] What object types in the digital heritage collections of your institution are accessible online as digitally reproduced or born digital objects?

(You can select more than one of the broad object classes presented here below. Please note: in the next screen(s) more detailed information about accessibility will be asked per object type.)

(Please use the table below to characterize the digital heritage collections of your institution. More than one object type may be selected.)

[12a] In what quantities are the following objects - as digitally reproduced or born digital objects - accessible online?

(the estimated number of units in the entire digital collection; please fill in "0" if for certain object types, for whatever reason, there are no objects accessible online)

	Term ([digital] objects)	Estimated # of units in entire digital collection	Units (#=Number of)
ARCHIVAL RESOURCES			
	Archives : Government documents		# Records
	Archives : Government archives		# Archives
	Archives : Other archival records		# Records
	Archives : Other archives		# Archives
TEXT BASED RESOURCES			
	Books		
	Rare printed books		# Records
	Other printed books		# Records
	Electronic books (eBooks)		# Records
	Serials		
	Newspapers		# Issues
	Journals		# Issues
	Articles		# Records
	Other serials		# Issues
	Other text based resources		
	Medieval Manuscripts		# Records
	Other Manuscripts		# Records
	Microforms / Microfilms		# Records

	Other text based		# Records
VISUAL (2D) RESOURCES			
	Drawings		# Records
	Engravings / Prints		# Records
	Maps and ground plans		# Records
	Paintings		# Records
	Photographs		# Records
	Posters		# Records
	Sheet music		# Records
	Other visual resources		# Records
3D MOVABLE OBJECTS			
	3 Dimensional works of art		# Records
	Archaeological man-made objects		# Records
	Coins and medals		# Records
	Other 3 dimensional man-made objects		# Records
GEOGRAPHY BASED RESOURCES			
	Landscapes		# Records
	Archaeological sites		# Sites
	Monuments and buildings		# Records
	Other geography based resources		# Records
NATURAL RESOURCES			
	Natural inert specimens		# Records
	Natural living specimens		# Records
TIME BASED RESOURCES			
	Audio files: Music		# Records
	Audio files: Speech & other (excl. digital audio books; incl. oral history		# Records

	files)		
	Digital audio books		# Records
	Film (Cinematic)		# Records
	Video recordings		# Records
	Other time based resources		# Records
DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]			
	Databases (containing cultural heritage metadata)		# Databases
	Digital (3D) designs or reconstructions of objects and buildings		# Records
	Digital art objects		# Records
	Digital research files (incl. GIS files)		# Records
	Games		# Records
	Software (customized)		# Records
	Websites (and parts of websites)		# Records
	Other digital interactive resources		# Records

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[13] What kind of descriptive metadata/cataloguing data are you providing online?

(Please check all the options that apply to your institution.)

- No metadata at all
- Metadata about items in online accessible digital collections
- Metadata about items in all digital collections, including offline digital collections
- Metadata about items that are not digitally available
- Other metadata (please specify below)

Other kinds of metadata are:

(Free text description of metadata that are provided online.)

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[14] Please indicate the estimated percentage of all the digital objects you have (digitally reproduced and/or born digital) that are and/or will be accessible through mentioned access options:

	% of digital objects currently accessible	% of digital objects accessible 2 years from now
Offline		
Institutional website		
National aggregator		
Thematic aggregator		
Europeana		
Memory of the World		
Wikipedia		
Institutional API		
3rd party API		
Social media platforms like Flickr or Facebook		
Other access options (please specify)		

Other access options are:

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

SECTION 4/6 - CONDITIONS OF USE

[15] Under what conditions do you provide access to/allow usage of your digital collections? (Please, check everything that is applicable for at least a part of your online digital collection on one or more platforms.)

	Objects	Metadata
Free restricted access (for instance with a login)		
Paid restricted access (for instance with a login)		
Restricted access for certain countries (geo-blocking)		
Non-restricted access for viewing		
Paid access for downloads, sharing & reuse		
Free non-commercial downloads, sharing & reuse		
Free commercial downloads, sharing & reuse		

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

SECTION 5/6 - MONITORING

[16] Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?

Please choose only one of the following:

- Yes
- No
- Do not know

(In order to be able to answer this question with Yes any manner of measurement will suffice.)

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[17] If Yes, which of the following methods does your organisation use to measure the number of times digital metadata and/or digital objects are being accessed?

[Only answer this question if the following conditions are met: Answer was 'Yes' at question #16 ("Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?")]

Please choose all that apply:

- Web statistics
- Database statistics (if not included in Web statistics)
- User studies
- Other (please specify):

Other methods of measuring the number of times the digital heritage collections of our institution are being accessed are:

(Please indicate all other ways in which the accessing of digital metadata and objects is measured. Do not mention specific software tools here. You can give the details about these in the next question.)

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[18] What tool(s) is/are being used in your organisation for measuring the number of times digital metadata and/or digital heritage objects are being accessed?

[Only answer this question if the following conditions are met: Answer was 'Yes' at question #16 ("Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?")]

Monitoring tools are:

(Please name all tools being used in your organisation for measuring the number of times digital metadata and/or digital objects are being accessed. If possible add any remarks that would help us better understand the practice within your institution.)

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[19] How often do you analyse the results of your access/use monitoring activities? Please give an explanation of your answer:

[Only answer this question if the following conditions are met: Answer was 'Yes' at question #16 ("Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?")]

- On a weekly basis
- On a monthly basis
- On a yearly basis
- Other (please specify):

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[20] In what way does your organisation make use of the outcomes of access/use data analysis?

[Only answer this question if the following conditions are met: Answer was 'Yes' at question #16 ("Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users?")]

The analysis of access/use data is used for:

(Please explain how your organisation valorises the collected data.)

Key Performance Indicators (KPIs) are "commonly used by an organisation to evaluate its success or the success of a particular activity in which it is engaged." [Wikipedia] In this context we are interested in how the monitoring of access and use in various cultural heritage institutions is related to assessing the successfulness of the organisation in fulfilling its mission.

[21a] Does your organisation apply Key Performance Indicators based on your monitoring [access/use] data?

- Yes
- No
- Do not know

[21b] If Yes, what KPI's does your organisation apply?

(Please briefly mention the KPIs that are used within your organisation.)

[21c] If Yes, *how* does your organisation report on these KPIs?

(Please describe the way in which KPIs are actually being used in showing the successfulness of your organisation in fulfilling its mission. Are KPIs mentioned in the institution's annual report? Are KPIs published on the institutional website? Etc.)

Evaluative remarks to this question:

(Please include any comments that would help us improve the question.)

[22] Please include any information that was not asked for above and that you think is relevant for understanding the nature of the access/use monitoring activities of your organisation:

SECTION 6/6 - FOLLOW-UP SURVEY

[23] The ENUMERATE Team plans to do additional research into monitoring web traffic for cultural heritage institutions (data analytics). Would your institution be interested in participating in follow-up research on data analytics?

- Yes
- No

If Yes, please provide us with a name and email address of the person in your institution who will be ENUMERATE's contact in analysing user statistics (if not yourself).

[24] Please provide us with a name and email address of the person in your institution who may be contacted in follow-up research on data analytics/user statistics (if not yourself).

(The full name and title of the person that can be approached for the user statistics research.)

[25] E-mail address:

(The primary e-mail address of the person to be approached for the user statistics research.)

E001 How long did it take to fill out the questionnaire?

E002 Did you have to look up things before filling out certain questions?

Thank you for collaborating with us in this Survey!

More information on the ENUMERATE Thematic Network is available on: www.enumerate.eu.

SIZE

ENUMERATE SURVEY ON THE SIZE OF DIGITAL COLLECTIONS

In this experimental thematic survey the aim is to develop better procedures for measuring the size of the digital heritage collections and the rate at which they are growing, both through the digitisation of analogue collections and the acquisition of new (born digital) materials.

We have developed three alternative approaches to measure the size of your collections. We ask each institution to fill out the questionnaire for two approaches, which we have selected at random for you.

It is important that the same person will try out the alternative approaches. At the end of the questionnaire we will ask you a few questions about these approaches: which one was most easy to handle (in terms of effort and time), etc.

Since the questionnaire aims to improve the methodology of monitoring the access and use of digital heritage collections, some questions have an accompanying field for evaluative remarks. Please help us with any comments that may be useful.

The time needed to fill in the questionnaire will depend highly on the availability of management information about your digital collections. If this information is in place, we estimate that answering the questions will take about 30 to 45 minutes of your time.

We hope you will be able to complete the questionnaire by **March 20th.** The questionnaire is designed in such a way that you can pause and return at any time to continue.

If you have any further questions, please contact your national coordinator, or:

Best Regards, the ENUMERATE Team

Natasha Stroeker: enumerate@panteia.nl

SECTION 1/5 - ORGANISATIONAL INFORMATION

[1] Name of institution / organisation *:

(Please provide the full, legal name of the company or institution about which you are answering the questions in this survey.)

[2] Type / Domain of institution / organisation *

Please specify the primary heading under which your company or institution classifies itself:

- National archive
- Other archive/Records office
- Audio-visual or broadcasting institution
- Film institute
- Museum of art
- Museum of archaeology, history
- Museum of natural history and natural science
- Museum of science and technology
- Museum of ethnography and anthropology
- Other type of museum
- National library
- Higher education library
- Special or other type of library
- Institution for Monument Care

[3] Your name:

(The full name and title of the person completing this survey.)

[4] Your telephone number and/or Skype contact details:

(The primary phone number (e.g. +44 0123456789) and/or the Skype details of the person completing this survey.)

[5] Your role in the institution:

(The position (job title) of the person completing this survey.)

[6] Your e-mail address:

(The primary e-mail address of the person completing this survey.)

[7] What is your institution's annual revenue budget?

Please choose only one of the following:

- < 10,000 €
- 10,000-50,000 €
- 50,000-100,000 €
- 100,000-500,000 €
- 500,000-1,000,000 €
- 1,000,000-10,000,000 €
- >10,000,000 €

(Provide the annual budget for the *entire* institution, for instance as indicated in the last set of published accounts. The budget may include government funding, project funding, revenues from the organisation's primary or commercial activities, etc. If choosing between two of the options is difficult (like when the budget is exactly 50,000 €), please choose the option with the lower estimate.)

[8] Is curatorial care for the collections of your institution part of its mission?

- Yes
- No

(Answer this question with 'No' if your institution does not hold heritage collections or if you have collections (for example of books, films, music) to be lent by or sold to contemporary users without the explicit task of safeguarding the collections for future generations).

[9] Does your organisation have digital collections or is it currently involved in collection digitisation and digital born objects collecting activities?

- Yes
- No

(Only answer this question with 'no' if your institution does not have any collections of digital materials (including collections of digital born heritage materials), is not involved in any digitisation activities as to its heritage collections and does not have any established plans to start developing its digital collection in the near future/in 2013.)

SECTION 2/5 - COLLECTION REGISTRATION SYSTEM APPROACH

In this fairly straightforward approach the aim is to measure the size of your digital collections based on the quantities of descriptive metadata records in your Collections Registration System(s): How many objects are registered in the CRS? How many of these are available as digital born and digitally reproduced objects? (Everything that is not in the CRS need not be considered here.)

Before we start it is important to clearly mark the difference between (*item level) descriptive metadata* and *digital born and digitally reproduced objects*:

- **(item level) descriptive metadata**: the bibliographic and descriptive metadata needed to include item level records in your (online) catalogue.
- **digital born and digitally reproduced objects**: *digital surrogates,* created as a result of converting analogue materials to digital form (digitisation), and *born digital objects*, for which there has never been and is never intended to be an analogue equivalent.

In the first part of this section we ask you to briefly describe the Collection Registration System(s) in use in your organisation.

[10] What is the name of the CRS of your institution?

(The name under which your CRS is commonly known. If several CRSs are in use, please specify the names of all CRSs.)

[11] If your CRS is online available, what is the URL of the online access point?

(Please specify the URL for the online available CRS(s) of your organisation. Leave blank if is there is no such access point.)

O not available online

[13] What percentage of your collections is *not* registered in your CRS?

(Please estimate the percentage of your collections that has not been registered in descriptive metadata records.)

[14] What software runs the CRS?

(Please specify the CRS software. If several software solutions are in use, please specify all.)

[15] What is the total number of descriptive metadata records in your CRS?

(If several CRSs are in use please give the sum total of all metadata records in all CRSs.)

[16] What classification system or list of object types is in use to classify the objects recorded in your CRS?

(Think of standard lists like the Art & Architecture Thesaurus (AAT) Objects Facet, etc.)

[17] What is the total # of Gigabytes/Terabytes/Petabytes of your digital heritage collections as recorded in your CRS?

(This includes both the storage size of metadata and the corresponding digitally reproduced or born digital objects, in all available qualities (master files, derivatives, etc.). Please clearly indicate the unit of measurement (Gigabyte/Terabyte/Petabyte) here. If several CRSs are in use please give the sum total for all CRSs.)

[18] Please specify the quantities of the object types that are represented in your CRS

(Please start with the most numerous object types. Leave the remaining rows empty if you have only a small number of object types recorded in your CRS).

	Object type (e.g. "Photographs")	# (=number of) descriptive metadata records	# of digitally reproduced or born digital objects
1.	<e.g. manuscripts>	<e.g. 60.000>	<e.g. 10.000>
2.	...	...	...
3.	...	...	...

SECTION 3/5 - DETAILED MEASURING OF DIGITAL SUB-COLLECTIONS APPROACH (2)

In this second approach the aim is to measure the size of your digital heritage collections through a detailed quantification, starting from a pre-defined table of digitally reproduced and born digital objects.

Again it is important to keep in mind the difference between item level metadata records and digital born and digitally reproduced objects. If we ask you to specify the number of records, this concerns the records with related *digital born and/or digitally reproduced objects.*

[19] Please indicate the size of your digital heritage collections per object type (columns A and B) and assess the need to digitise for objects in your *entire* analogue collections (columns C and D).

(Use the table below to quantify the digital heritage collections of your institution.)

	Term ([digital] objects)	Units (#=Number of) [#Records means: # Metadata records in CRS]	Estimated # of units in entire digital collection (A)	Estimated % born digital of units in entire digital collection (B)	Estimated % of *entire* analogue collection still to be digitized (C)	Estimated % of *entire* analogue collection no need to be digitized (D)
ARCHIVAL RESOURCES						
	Archives : Government documents	# Records				
	Archives : Government archives	# Archives				
	Archives : Other archival records	# Records				
	Archives : Other archives	# Archives				
TEXT BASED RESOURCES						
Books						
	Rare printed books	# Records				
	Other printed books	# Records				
	Electronic books (eBooks)	# Records				

Serials						
	Newspapers	# Issues				
	Journals	# Issues				
	Articles	# Records				
	Other serials	# Issues				
Other text based resources						
	Medieval Manuscripts	# Records				
	Other Manuscripts	# Records				
	Microforms / Microfilms	# Records				
	Other text based	# Records				
VISUAL (2D) RESOURCES						
	Drawings	# Records				
	Engravings / Prints	# Records				
	Maps and ground plans	# Records				
	Paintings	# Records				
	Photographs	# Records				
	Posters	# Records				
	Sheet music	# Records				
	Other visual resources	# Records				
3D MOVABLE OBJECTS						

	3 Dimensional works of art	# Records				
	Archaeological man-made objects	# Records				
	Coins and medals	# Records				
	Other 3 dimensional man-made objects	# Records				
GEOGRAPHY BASED RESOURCES						
	Landscapes	# Records				
	Archeological sites	# Sites				
	Monuments and buildings	# Records				
	Other geography based resources	# Records				
NATURAL RESOURCES						
	Natural inert specimens	# Records				
	Natural living specimens	# Records				
TIME BASED RESOURCES						
	Audio files: Music	# Records				
	Audio files: Speech & other (excl. digital audio books; incl.	# Records				

	oral history files)					
	Digital audio books	# Records				
	Film (Cinematic)	# Records				
	Video recordings	# Records				
	Other time based resources	# Records				
DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]						
	Databases (containing cultural heritage metadata)	# Records				
	Digital (3D) designs or reconstructions of objects and buildings	# Records				
	Digital art objects	# Records				
	Digital research files (incl. GIS files)	# Records				
	Games	# Records				
	Software (customised)	# Records				
	Websites (and parts of websites)	# Records				
	Other digital interactive	# Records				

	resources					

SECTION 4/5 - IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

The third approach chosen here is based on estimates of the size of your digital heritage collections and the number of objects that still need to be digitised. This is supplemented by two questions about the born digital materials collected by your institution.

Again it is important to keep in mind the difference between *(item level) metadata records* and *digital born and digitally reproduced objects*. (See above.)

[20a] Estimate the percentage of your heritage collections that <u>has already been digitally reproduced</u>:

… %

(Digitally reproduced should be discerned from being catalogued in digital (textual) metadata records. A digital reproduction is a digital surrogate of the original analogue object. An object that has been catalogued in a digital (textual) metadata record is not -digitally reproduced- in the definition we use here. Please give a global estimate (percentage) <u>over all (analogue) heritage collections</u> in your institution.)

[20b] Please specify the object types and percentages that were included in the previous answer [20a].

In the table below:

- (A) Choose all object types that apply, in other words: which object types have been digitally reproduced?
- (B) Please estimate how these digitally reproduced objects are distributed in percentages amongst the different object types.

(Example: if 30% of your heritage collections has been digitally reproduced and your institution has three different types of objects in the digital collections, we ask you to divide this 30% over the three object types, e.g. 10% - 10% - 10% OR 5% - 15% - 10% etc.)

[NOTE: Table will not be presented as it is presented here. Preferably the table can be folded out, starting from the high level object classes.]

Which apply (A)	Term ([digital] objects)	Estimated % digitally reproduced (B)
ARCHIVAL RESOURCES		
◻	Archives : Government documents	
◻	Archives : Other archival records	
TEXT BASED RESOURCES		
Books		
◻	Rare printed books	
◻	Other printed books	
Serials		
◻	Newspapers	
◻	Journals	
◻	Other serials	

Other text based resources		
◘	Medieval Manuscripts	
◘	Other Manuscripts	
◘	Microforms / Microfilms	
◘	Other text based	
VISUAL (2D) RESOURCES		
◘	Drawings	
◘	Engravings / Prints	
◘	Maps and ground plans	
◘	Paintings	
◘	Photographs	
◘	Posters	
◘	Sheet music	
◘	Other visual (2D) resources	
3D MOVABLE OBJECTS		
◘	3 Dimensional works of art	
◘	Other (3D) man-made objects	
GEOGRAPHY BASED RESOURCES		
◘	Landscapes	
◘	Archeological sites	
◘	Monuments and buildings	
◘	Other geography based resources	
NATURAL RESOURCES		
◘	Natural inert specimens	

▫	Natural living specimens	
TIME BASED RESOURCES		
▫	Audio files: Music	
▫	Audio files: Speech & other	
▫	Film (cinematic)	
▫	Video recordings	
▫	Other time based resources	
▫	Other (please specify below)	

Other object types (not listed above) are:

(Free text description of digitised analogue objects, not listed.)

[21a] Estimate the percentage of your heritage collections that <u>still needs to be digitally reproduced</u>:

… %

(This is 100% - all of your collections - minus the percentage of your collections that has already been digitised, minus the percentage of your collections for which there is no need to be digitally reproduced. Digitally reproduced should be discerned from being catalogued in a digital (textual) metadata record. A digital reproduction is a digital surrogate of the original analogue object. An object that has been catalogued in a digital (textual) metadata record is not "digitally reproduced" in the definition we use here. Please give a global estimate (percentage) over all heritage collections in your institution.)

[21b] Please specify the object types and percentages that were included in the previous answer [21a].

In the table below:

- (A) Choose all object types that apply, in other words: which object types have been digitally reproduced?
- (B) Please estimate how these analogue objects that still need to be digitally reproduced are distributed in percentages amongst the different object types.

(EXAMPLE: if 70% of your heritage collections still needs to be digitally reproduced and you have three different object types in you collections, we ask you to divide this 70% over the three object types, e.g. 0% - 30% - 10% OR 20% - 15% - 35% etc.)

[NOTE: Table will not be presented as it is presented here. Preferably the table can be folded out, starting from the high level object classes.]

Which apply (A)	Term ([digital] objects)	Estimated % digitally reproduced (B)
ARCHIVAL RESOURCES		
▫	Archives : Government documents	

◘	Archives : Other archival records	
TEXT BASED RESOURCES		
Books		
◘	Rare printed books	
◘	Other printed books	
Serials		
◘	Newspapers	
◘	Journals	
◘	Other serials	
Other text based resources		
◘	Medieval Manuscripts	
◘	Other Manuscripts	
◘	Microforms / Microfilms	
◘	Other text based	
VISUAL (2D) RESOURCES		
◘	Drawings	
◘	Engravings / Prints	
◘	Maps and ground plans	
◘	Paintings	
◘	Photographs	
◘	Posters	
◘	Sheet music	
◘	Other visual (2D) resources	
3D MOVABLE OBJECTS		

◻	3 Dimensional works of art	
◻	Other (3D) man-made objects	
GEOGRAPHY BASED RESOURCES		
◻	Landscapes	
◻	Archeological sites	
◻	Monuments and buildings	
◻	Other geography based resources	
NATURAL RESOURCES		
◻	Natural inert specimens	
◻	Natural living specimens	
TIME BASED RESOURCES		
◻	Audio files: Music	
◻	Audio files: Speech & other	
◻	Film (cinematic)	
◻	Video recordings	
◻	Other time based resources	
◻	Other (please specify below)	

Other object types (not listed above) are:

(Free text description of digitised analogue objects, not listed.)

[22a] Does your organisation collect born digital heritage materials?

Please choose only one of the following:

- Yes
- No
- Do not know

Answer this question with ´Yes´ if your institution collects any kind of born digital heritage (i.e. software, digital documents, digital art, harvested web content, etc.).

[22b] If Yes, what types of born digital heritage?

Please specify the object types that were included in the previous answer:

Which apply (A)	Term ([digital] objects)
ARCHIVAL RESOURCES	
¤	Archives : Government documents
¤	Archives : Other archival records
TEXT BASED RESOURCES	
¤	Electronic books (eBooks)
¤	e-Newspapers
¤	e-Journals
¤	Other e-Serials
¤	Other digital born text documents
VISUAL (2D) RESOURCES	
¤	Digital maps and ground plans
¤	Photographs
TIME BASED RESOURCES	
¤	Audio files: Music
¤	Audio files: Speech & other
¤	Film
¤	Video recordings
DIGITAL INTERACTIVE RESOURCES	
¤	Databases (containing cultural heritage metadata)
¤	Digital (3D) designs or reconstructions of objects and buildings
¤	Digital art objects (incl. Internet art)
¤	Digital research files (incl. GIS files)
¤	Games

☐	Software (customised)
☐	Websites (incl. blogs, tweets, widgets, wikis)
☐	Other (please specify below)

Other object types (not listed above) are:

(Free text description of born digital objects, not listed.)

SECTION 5/5 - EVALUATIVE REMARKS

These questions were different for each of the three options. Each respondent answers the evaluative remarks for two of the three approaches that are selected. Here approach 1 and 3 are selected.

In this final section we would like to know about your experiences in filling out the questionnaire. Therefore we kindly ask you how time consuming and 'easy' or 'hard' it was to fill out the two approaches tried out above and which one provides the best insight into the size of the (digital) collections of your institution.

[23] How much time / effort was needed to fill out the questions for the two different approaches?

APPROACH	***# Minutes***
COLLECTION REGISTRATION SYSTEM APPROACH	
IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH	

(Please estimate the number of minutes spent on each of the approaches. If information collected in either one of the approaches was re-used in one or both the other approaches, please try to account for that by distributing this time proportionally.)

[24] How difficult/hard was it to collect the necessary information to answer the questions in each of the two approaches?

APPROACH	**1**	**2**	**3**	**4**	**5**
COLLECTION REGISTRATION SYSTEM APPROACH					
IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH					

(Using a 5-points scale - where 1 equals "very difficult/hard" to 5 "not at all difficult/very easy" - please select only one number per row. Again, please try to account for the reuse of information, as in the previous question.)

[25a] Which one of the approaches would you prefer to fill out in future issues of the ENUMERATE Core Survey?

- COLLECTION REGISTRATION SYSTEM APPROACH
- IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

(Please select only one of the alternative approaches.)

[25b] Why do you prefer the approach selected?

(Please explain you answer to question 25a.)

[26a] Which one of the alternative approaches will offer in your opinion the most faithful insight into the actual size of the (digital) collection of your institution?

- COLLECTION REGISTRATION SYSTEM APPROACH
- IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

(Please select only one of the alternative approaches.)

[26b] And why?

(Please explain you answer to question 26a.)

[27] Please include any comments that would help the community of museums, libraries, archives and caretakers of monuments to improve the methodology of measuring the size and growth of digital heritage collections:

Thank you for collaborating with us in this Survey!

More information on the ENUMERATE Thematic Network is available on: www.enumerate.eu.

COST

In this thematic test survey the aim is to get a grasp of the cost of digital collections. Specialists in selected institutions are asked to consider the costs of acquiring and valorising one specific digital collection. Costs are divided into incidental (upfront) costs and structural (ongoing) costs.

- Incidental costs are defined as the costs having to do with the initial creation or acquisition of a digital collection.
- Structural costs are the (annual) costs needed for the ongoing maintenance, enhancement and preservation of a digital collection

The survey consists of 14 questions. Since the questionnaire aims to improve the methodology of monitoring cost, some questions have an accompanying field for evaluative remarks. Please help us with any comments that may be useful.

We estimate that answering the questions will take about 15-45 minutes of your time.

We hope you will be able to complete the questionnaire by April 16. The questionnaire is designed in such a way that you can pause and return at any time to continue. After you have finished and submitted the questionnaire, we may possibly contact you in order to resolve ambiguities.

If you have any further questions, please contact your national coordinator, or:

Natasha Stroeker: enumerate@panteia.nl

Best Regards, the ENUMERATE Team.

SECTION 1/2 - ORGANISATIONAL INFORMATION / INSTITUTIONAL PERSPECTIVE

[1] Name of institution / organisation *:

(Please provide the full, legal name of the company or institution about which you are answering the questions in this survey.)

[2] Type / Domain of institution / organisation *

Please specify the primary heading under which your company or institution classifies itself:

- National archive
- Other archive/Records office
- Audio-visual or broadcasting institution
- Film institute
- Museum of art
- Museum of archaeology, history
- Museum of natural history and natural science
- Museum of science and technology
- Museum of ethnography and anthropology
- Other type of museum
- National library
- Higher education library
- Special or other type of library
- Institution for Monument Care

[3] Your name:

(The full name and title of the person completing this survey.)

[4] Your telephone number:

(The primary phone number (e.g. +44 0123456789) of the person completing this survey.)

[5] Your role in the institution:

(The position (job title) of the person completing this survey.)

[6] Your e-mail address:

(The primary e-mail address of the person completing this survey.)

[7] What is your institution's annual revenue budget?

Please choose only one of the following:

- < 10,000 €
- 10,000-50,000 €

- 50,000-100,000 €
- 100,000-500,000 €
- 500,000-1,000,000 €
- 1,000,000-10,000,000 €
- >10,000,000 €

(Provide the annual budget for the *entire* institution, for instance as indicated in the last set of published accounts. The budget may include government funding, project funding, revenues from the organisation's primary or commercial activities, etc. If choosing between two of the options is difficult (like when the budget is exactly 50,000 €), please choose the option with the lower estimate.)

[8] Does the most recent annual report of your institution explicitly mention costs related to your institution's digital collections?

- Yes
- No
- Do not know

[9] If you answered Yes in the previous question what does this most recent annual report of your institution mention:

- Growth of digital collections
- FTEs involved in creating and managing digital collections
- Budget related to creating and managing digital collections
- Growth of digital storage (in Terabyte)
- Other (please specify below)

Other costs mentioned in the most recent annual report of your institution are:

SECTION 2/2 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION

In this section the focus will be on one specific major digital heritage collection within your institution. Preferably this is a prominent digital collection that has been publicly available (online) for some time. The collection may contain one or more types of digital objects (e.g. digitised paintings, digitised rare books, digital 3D reconstructions of buildings, digital photographs, etc.) and will have the associated metadata recorded in a collection registration system. If your institution does not have a digital heritage collection exactly matching the description above, please select the collection that comes most close.

[10] Please give a name or description of the selected collection:

[11] If the collection is made accessible online, please give a reference to it (URL):

[12] What quantities of digital object types are in the selected collection?

(Please use the table below to characterize the selected collection. More than one object type may be selected. Where the unit is the number of records, a 'record' is a record in your collection registration system, i.e. the container of bibliographic and descriptive metadata.)

	Term ([digital] objects)	Estimated # of units in entire digital collection	Units (#=Number of)
ARCHIVAL RESOURCES			
	Government documents		# Records
	Government archives		# Archives
	Other archival records		# Records
	Other archives		# Archives
TEXT BASED RESOURCES			
	Books		
	Rare printed books		# Records
	Other printed books		# Records
	Electronic books (eBooks)		# Records
	Serials		
	Newspapers		# Issues
	Journals		# Issues

	Articles		# Records
	Other serials		# Issues
	Other text based resources		
	Medieval Manuscripts		# Records
	Other Manuscripts		# Records
	Microforms / Microfilms		# Records
	Other text based		# Records
VISUAL (2D) RESOURCES			
	Drawings		# Records
	Engravings / Prints		# Records
	Maps and ground plans		# Records
	Paintings		# Records
	Photographs		# Records
	Posters		# Records
	Sheet music		# Records
	Other visual resources		# Records
3D MOVABLE OBJECTS			
	3 Dimensional works of art		# Records
	Archaeological man-made objects		# Records
	Coins and medals		# Records
	Other 3 dimensional man-made objects		# Records
GEOGRAPHY BASED RESOURCES			
	Landscapes		# Records
	Archaeological sites		# Sites

	Monuments and buildings		# Records
	Other geography based resources		# Records
NATURAL RESOURCES			
	Natural inert specimens		# Records
	Natural living specimens		# Records
TIME BASED RESOURCES			
	Audio files: Music		# Records
	Audio files: Speech & other (excl. digital audio books; incl. oral history files)		# Records
	Digital audio books		# Records
	Film (Cinematic)		# Records
	Video recordings		# Records
	Other time based resources		# Records
DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]			
	Databases (containing cultural heritage metadata)		# Records
	Digital (3D) designs or reconstructions of objects and buildings		# Records
	Digital art objects		# Records
	Digital research files (incl. GIS files)		# Records
	Games		# Records
	Software (customized)		# Records

	Websites (and parts of websites)		# Records
	Other digital interactive resources		# Records

[13] Activity based costing: determining Incidental costs

We now invite you to estimate how much your institution has spent (in Euros) for *the first time creation or acquisition* (incidental costs) of the selected digital collection (including the cost of staff time) for each of the following activities, wherever these occurred in the institution or have been outsourced.

Leave blank the items that you cannot estimate. Enter 0 if a cost item is not applicable. If you miss any items in the table, please help us and mention these under *Other costs*.

(The time period concerned can be indicated below the table.)

Cost item	In-house #EURO	Outsourced #EURO
Project management		
Selection of material for digitisation		
Acquisition of digital born material		
Logistics (shipment of collection for digitisation, etc)		
Analogue-Digital conversion (including all technical and staff costs associated with the act of preparing materials for scanning, the scanning itself, and quality control):		
Copyright clearance		
Metadata creation and/or enhancement		
Collection of user generated content (UGC)		
Web design and/or software development		
Other costs, please specify: […]		

(In the #EURO cells enter a whole number without a currency symbol or commas.)

Please indicate the time span (in months) during which this budget was spent:

(For example: if your institution has created the selected digital collection between August 2002 and September 2004, fill in 25 months, etc.)

Evaluative remarks to this question:

(Please indicate if the above mentioned activities represent parts of your workflow. If Yes, were you able to determine faithfully the costs of the various activities? If No, were you able to break down your workflow in alternative activities and accessory costs?)

[14] Activity based costing: determining Structural costs

For ***the past fiscal year***, please estimate the structural costs for the selected digital collection. Fill in the form for each cost item that can be estimated, wherever it occurs in the institution or is outsourced.

Leave blank the items that you cannot estimate. Enter 0 if a cost item is not applicable. If you miss any items in the table, please help us and mention these under *Other costs*.

Cost item	In-house #EURO	Outsourced #EURO
Project management		
Archiving (storage, including backups)		
Activities concerning the (long-term) preservation of the digital collection (storage costs excluded).		
Licensing		
Costs related to giving access to the digital collection (e.g. maintenance web servers and web services)		
User outreach and support (including all staff time for efforts to promote the use of the collections)		
Usage analysis (including surveys, interviews, and other activities)		
Editorial (including content selection and updating)		
Other costs - please specify: …		

(In the #EURO cells enter a whole number without a currency symbol or commas.)

Evaluative remarks to this question:

(Please indicate if the above mentioned activities represent parts of your workflow. If Yes, were you able to actually determine the costs of the various activities? If No, were you able to break down your workflow in alternative activities and accessory costs?)

Thank you for collaborating with us in this Survey!

More information on the ENUMERATE Thematic Network is available on: … Etc.

Appendix 3 Preferred and most accurate approach - SIZE

Reasons for preferred approach

Approach 1: Collection registration system approach

- *This questionnaire seems to be a very theoretic approach. In a large museum, which also comprises not only collections but also a large library and two archives and other collections such a detailed questionnaire makes no sense at all. It is much too time consuming to answer these questions. You should start again with much more general questions.*
- *It is the lighter one for our organisation (Since not perfect since we are not a museum neither an archive and do use a different vocabulary and tools).*
- *It is more coherent with the data we have. We do not have detailed estimates of the number of articles in newspapers where we only have an estimate of the number of issues. The detailed digital sub-collections survey is too detailed for resources which are still analogue and cannot be measured automatically. Furthermore, the biggest problem in the second measurement is the fact that there is no distinction between whether you actually have objects of that type in analogue form or whether you don't want to digitize them.*
- *The approach B implies confusing questions. Namely it is confusing to which reference figures the percentages are referring.*
- *More adapted to the characteristics of archives*
- *Because the questions mirror the structure of our collection more appropriately*
- *It is easier, if you have useful statistics of CRS.*
- *Easier*
- *Coordination of digitalization is easier because we are part of union catalogue*
- *Because we make statistics, that fits into questions in this part of questionnaire.*
- *Both approaches are needed to get a overall view of the situation. The questionnaire could preferable be shorter. The time span for answering should be longer because the information has to be gathered from many departments in bigger organisations.*
- *We use similar system for statistic.*
- *Easier to fill*
- *Difficult to estimate percentages in second approach.*
- *The variety of objects (in more than ten collections!) that the museum safeguards makes it difficult to estimate the digital materials. And digitalization is not definite clearly enough: is the scan of a cover of a manuscript digitalization or only its full digitalization?*

Approach 2: Detailed measuring of digital sub-collections approach

- *CRS is not currently applicable to the majority of our collections*
- *Gives more information and data about the digitised collections*
- *Because we are going to digitize such a small percentage of our total collections, the numbers that had to be divided per object type was closer to zero in the improved survey.*
- *Because it is simple way to measure the objects and what of those objects has already digitally reproduced. For example, we have a painting collection of modest size, about 6500 objects. For me it is important to know how many out of these 6500 objects has already been digitally reproduced not so much to know how it is related to digitization of entire collection. On some occasions it is important to get detailed insight of entire collection but in my everyday work it has little use.*
- *It is part of our normal measurement work and easy to pick up from databases.*
- *Because we don't have born digitals, it goes faster. More understandable, more clearly than the other approach*

Approach 3: Improved ENUMERATE Core Survey questions approach

- *Because I can be followed the size of the collection still need to be digitized also in the future and divides the percentages per objects that have to be digitalised.*
- *It is less complicated.*
- *All the data is provided by our data base*
- *The digital collections we hold are not very extensive and there is also uncertainty about the extent of overall holdings.*
- *I think it is easier to handle.*
- *Easier to understand (but perhaps this is because I went through the other approach before?)*
- *We do not use CRS*
- *I prefer check boxes because of standardizing effects.*

Reasons for most accurate approach

Approach 1: Collection registration system approach

- *None of these two approaches really help in the specific case of our organisation since we are not a museum neither an archive and are running our own system. We were obliged to make compromise in order to fit with the fields of this survey and to be able to fill it within a reasonable duration.*
- *Gives more realistic overview of our digital collections*
- *The basis of our treatment are funds and collections.*
- *As said before: It reflects the structure of our collection in a better way*
- *In our case the first approach is more faithful, because all digitized objects are in our CRS, which allows getting actual size of objects.*
- *More precise*
- *All data is in one place*
- *The numbers are accurate.*
- *Better overview*
- *Estimates offered in second approach may be less accurate.*

Approach 2: Detailed measuring of digital sub-collections approach

- *Cover the most complete range of resources and collections within the library*
- *Because it includes the combined information, of digitalised collections in relation to analogue collections.*
- *Once the objects are digital and hosted by yourself, it makes sense to measure them in detail. However, this doesn't work for analogue material or licensed content hosted by third parties.*
- *The concrete numbers could show how much work still has to be done. It provokes important questions about in which direction the archives will develop in the future*
- *It's our normal way to measure it*
- *It conforms more to the structure of our collection*

Approach 3: Improved ENUMERATE Core Survey questions approach

- *It gives detailed insight of the digitization process.*
- *Seems more useful for our purposes*
- *I am working in a museum. We have a database in which there are three kinds of material: items, pictures, drawings. I have the feeling that the questions don't show very clearly the situation of our*

museum.

- *I think/ hope so*
- *Actually results for both approaches should be about equally meaningful. But there is no tick box for this.*
- *You can see the numbers of artefacts*
- *Even if it's difficult for museums with more than one or two different types of objects to estimate the quote of digitalization, it is clearly a more faithful insight into the variety of digitalzed objects. Because it also shows the difficulty of digitalization: documents with one or two pages are much easier to digitalize than a three-dimensional object.*