



Grant Agreement 270939

ENUMERATE

Conceptual Framework

Deliverable number	<i>D2.12</i>
Dissemination level	<i>PU</i>
Delivery date	<i>April 2014</i>
Status	<i>Final after Review</i>
Author(s)	<i>Gerhard Jan Nauta & Marco de Niet (DEN)</i>



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

Document history

Revision	Date	Author	Organisation	Description
V0.1	2014-01	Gerhard Jan Nauta Marco de Niet	DEN DEN	Draft
V0.2	2014-02	Gerhard Jan Nauta Marco de Niet	DEN DEN	Revised after feedback by CT
V1.0	2014-02	Gerhard Jan Nauta Marco de Niet	DEN DEN	Final after revision by Core Group
Reviewed by		Natasha Stroecker	Panteia	
		René Vogels	Panteia	
		Jesús Domínguez	DIGIBIS	
		Monika Hagedorn-Saupe	SPK	
		Nick Poole	CT	
		Gordon McKenna	CT	
V2.0	2014-04	Gerhard Jan Nauta	DEN	Revised after Project Review 18/3/14
V2.1	2014-04	Marco de Niet	DEN	Final editing

Contents

1	INTRODUCTION.....	4
2	LESSONS LEARNED	5
2.1	<i>Simultaneous surveying across Europe takes tremendous efforts</i>	6
2.2	<i>Europe-wide surveying requires compromising</i>	6
2.3	<i>Commitment is crucial.....</i>	7
2.4	<i>Customisation is needed to increase level of participation</i>	7
2.5	<i>No more comprehensive Thematic Surveys</i>	7
2.6	<i>Many institutions lack management data about digital heritage</i>	8
2.7	<i>Finding the right person to address is difficult and time consuming</i>	8
2.8	<i>Contact databases are rapidly becoming unreliable.....</i>	8
2.9	<i>Flexible deadlines increase levels of participation</i>	9
2.10	<i>Reporting on subsets of data deserve more attention</i>	9
2.11	<i>Cross-domain harmonisation deserves more attention</i>	9
2.12	<i>Measuring digital preservation has not yet reached a mature level.....</i>	9
2.13	<i>Conclusion.....</i>	10
3	FRAMEWORK (METHODOLOGY).....	11
3.1	<i>A definition of the cultural heritage domain in the context of this Framework.....</i>	11
3.2	<i>The size of the European cultural heritage domain</i>	11
3.3	<i>Baseline data collected in two subsequent Core Surveys</i>	15
3.4	<i>Target audience and reporting formats</i>	15
3.5	<i>Minimum set of (core) indicators for ‘high level’-surveys</i>	16
3.6	<i>More in-depth indicators for thematic surveys.....</i>	18
3.7	<i>Statistical standards applying to all indicators</i>	18
3.8	<i>A procedure for the assimilation of new indicators</i>	20
3.9	<i>Definitions and Tools.....</i>	21
3.10	<i>Liaising with other monitoring practices.....</i>	21
3.11	<i>Suggestions for the organisation of future data collection activities</i>	21
3.12	<i>Planning for future surveys</i>	24
4	RECOMMENDATIONS	25
4.1	<i>Recommendations for Cultural Heritage Institutions</i>	25
4.2	<i>Recommendations for Governments and Umbrella Organisations</i>	26
5	APPENDICES.....	29
5.1	<i>Core indicators for ‘high level’-surveys</i>	29
5.2	<i>Classification of heritage materials</i>	38
5.3	<i>The ENUMERATE Definitions</i>	41
5.4	<i>The ENUMERATE Core Survey 2 questionnaire.....</i>	49

1 Introduction

The objectives of the *ENUMERATE* Thematic Network (2011-2014) were to:

- Gather general statistical information and build a baseline of data on the state of digitisation in museums, archives, libraries and other memory institutions in Europe.
- Understand the progress that has been made in Europe since the *NUMERIC* survey was conducted.
- Achieve a situation where informed decisions about national and EU policies are possible.
- Help the memory institutions in measuring their progress in the field of digitising heritage materials and in making decisions of a practical nature.

This report is the final deliverable of the *ENUMERATE* Work Package 2, *Methodology*. This deliverable wraps up all the findings and conclusions reached during the work on the methodology for *ENUMERATE* and brings it together as a *Conceptual Framework for capturing statistical information about digital activities of European cultural heritage institutions*.

A 'conceptual framework' sets standards that guide the collection of data. It provides both a description of the elements that a field of study consists of - in this case that field is the domain of digital heritage, more specific the European memory institutions holding or planning to hold digital collections -, and an outline of the relationships between the elements in that field.¹

This *Framework* is intended to serve as a starting point for future research and data gathering to improve the understanding of the status of digital heritage collections and services through statistical data. It brings together knowledge acquired during the *ENUMERATE* project, including final versions of supporting tools that were created during the project.

The *ENUMERATE Conceptual Framework* provides:

- 1 A definition of the cultural heritage domain in the context of this *Framework*;
- 2 Information about the size of the European cultural heritage domain;
- 3 Baseline data collected in two subsequent *Core Surveys*;
- 4 Information about target audience and reporting formats;
- 5 Core indicators for 'high level' surveys;
- 6 More in-depth indicators for thematic surveys;
- 7 Statistical standards applying to all indicators;
- 8 A procedure for the assimilation of new indicators;
- 9 Definitions and Tools;
- 10 Information about liaising with other monitoring practices;
- 11 Suggestions for the organisation of future data collection activities;
- 12 Planning for future surveys.

Options to actually take the *ENUMERATE* framework forward are not discussed here. They are described in the *Business Plan and Sustainability Strategy* (deliverable D4.8).

The *Conceptual Framework* is presented in Chapter 3 of this document. Before we present the *Framework*, we will share some lessons learned from the work on the *ENUMERATE* methodology in Chapter 2. In many ways the modelling of the *Framework* will be derived from these lessons learned.

The following appendices have been attached:

- 4.1 Core indicators for 'high level' surveys;
- 4.2 Classification of heritage materials;
- 4.3 The *ENUMERATE* Definitions;
- 4.4 The *ENUMERATE Core Survey 2* questionnaire.

¹ ITU, *Measuring the WSIS Targets. A statistical framework* (2011), p.1.

2 Lessons Learned

In this chapter the main lessons learned during the work on the methodology in the *ENUMERATE* project are presented. The lessons are partly based on the project's experiences, partly on the feedback received from stakeholders (in- and outside of the project consortium) while running the *ENUMERATE* surveys. The main stakeholders of *ENUMERATE* Thematic Network (TN) were:

European Commission:

- Grant supplier to the *ENUMERATE* TN.

ENUMERATE TN Team:

- Initiators of the *ENUMERATE* TN;
- Development of *ENUMERATE* methodology (questionnaire, definitions, survey process, and time schedule);
- Preparation of 2 *Core Surveys* and 1 *Thematic Survey* (including survey tools);
- Coordination of the survey process;
- Communication activities.

ENUMERATE TN Consortium:

- Advising on all key aspects of the *ENUMERATE* project;
- Quality assurance;
- Communication activities.

ENUMERATE TN Advisory Group:

- Advising on *ENUMERATE* methodology and surveys.

Data Partner (Panteia):

- Advising on *ENUMERATE* methodology;
- Running the surveys (including collecting the data);
- Analysis of survey results;
- Reporting on survey outcomes.

National Coordinators:

- Deciding on level of involvement per country (Gold-Silver-Bronze scenario's);
- Support for questionnaire development (e.g. translations);
- Providing database of contacts/respondents;
- Supporting the survey process at the national levels;
- Reporting on survey outcomes at the national levels.

Cultural Heritage Institutions:

- Contribute to the surveys based on their own management and collection data.

Creating the methodology for the *ENUMERATE* surveys was an iterative process involving the stakeholders at different stages during the creation. Considering the diversity and heterogeneity of the group of stakeholders, it proved to be worthwhile to have many consultation and feedback opportunities to help refine and improve the methodology. This led to a better acceptance of and support for the survey. For example, the process for creating *ENUMERATE Core Survey 2* took the following elements into account:

- Feedback from individual respondents (memory institutions) to *Core Survey 1*;
- A formal review of *Core Survey 1* methodology by the Data Partner Panteia;
- Recommendations from the four specialist meetings organised between 2012 and 2013;
- Results from the official project reviews in 2012 and 2013;
- Lessons learned from three *Thematic Surveys* held in February-April 2013;
- Feedback from the EU Member States Expert Group in April 2013;
- A review by the Consortium partners on July 10th, Vienna and e-mail communications in the weeks thereafter;
- Intensive communications with the National Coordinators during the development of the online CS 2 questionnaire in various languages;
- Continuous fine tuning of questionnaires and methodology with Panteia and Digibis;
- Miscellaneous feedback and acuminate comments from various stakeholders, including the members of the Advisory Group.

From these diverse and dynamic processes we present here the most important lessons learned.

2.1 Simultaneous surveying across Europe takes tremendous efforts

The central characteristic in the approach of measuring the progress of digital cultural heritage in Europe was the harmonisation of measuring across all EU countries. To enable this, the *ENUMERATE* Team chose to run the surveys in limited time frames across all EU countries (supplemented with the EFTA countries).

A monolithic questionnaire was developed, which was translated and distributed simultaneously. The time the *ENUMERATE* Team had to spend on planning and producing translations of the Core Survey questionnaire was substantial. The team had to compromise on the quality of the translations. A thorough procedure would provide in quality checks, which might easily result in a doubling of efforts.

The main lesson we can learn from the project is that it takes tremendous efforts from a large group of people to achieve full synchronisation across all the countries.

For future surveys, we would like to recommend a more modular approach, to be more flexible and adaptable in collecting relevant data. This can be done, for instance, by tapping into national surveys, as run by some national governments. This may also diminish the need for central funding. However, a central place where the data can be collected, analysed and published will still be needed. *ENUMERATE* sees the *Europeana Pro* platform as a good vehicle to this end.

2.2 Europe-wide surveying requires compromising

Creating a methodology to measure the progress of digital heritage in Europe turned out to be a balancing act between statistical soundness and political and institutional idiosyncrasies. During the project we had to make quite a few compromises in the development of the methodology to keep cultural heritage institutions and national coordinators on board. For instance, the option to arrive at a fully random sample of institutions to be involved in the questionnaire had to be dropped, as some national coordinators wanted to have all major players in the country on board in the questionnaire, otherwise they would not participate.

2.3 Commitment is crucial

The *ENUMERATE* project was supported tremendously by a small group of people across Europe who truly believed in the usefulness of the work done in the project, for which there was hardly any alternative at hand. The commitment they showed to stay involved without any form of compensation, understanding that the project could make a crucial contribution to a complex but necessary task, was much appreciated.

It should be noted that one of the official EU reviewers clearly stated in a meeting that he considered the *ENUMERATE* project to be both under-resourced and under-staffed, considering the importance of the work done.

It should also be noted that the consortium, from 'day one', believed that the format chosen by the European Commission, a Thematic Network, was not the right vehicle for the project, especially with regards to the funding mechanism. All partners have done a lot of work without any financial compensation at all, in order to free funds for contracting a specialised partner to run the surveys and do the analysis.

The lesson learned is that setting up and running a framework for digital heritage statistics is a topic that is strongly believed in by a small group of professionals and that it takes a lot of effort to motivate and include a much wider group of people to stay involved in maintaining the framework.

2.4 Customisation is needed to increase level of participation

ENUMERATE did not have the means to pay institutions to provide their data. Their involvement was on a voluntary basis. Also the National Coordinators donated their time to the project for free, and could not be compensated. As top-down financial incentives were lacking, a demand driven approach was needed. It is expected to be the same for future survey efforts. *ENUMERATE* tried to make the benefits explicit for both the institutions, both in direct communication and through services like the online benchmarking tool.

Also some time-consuming questions were made optional in the survey. For the National Coordinators three scenarios (Gold-Silver-Bronze) were conceived that related to the time they could commit to the project in getting the survey distributed in their country. For the continuation of the framework, it is considered crucial to keep working on making incentives explicit and invest in methods to reduce the time as much as possible, without compromising the quality of the work too much.

2.5 No more comprehensive Thematic Surveys

In *ENUMERATE* three surveys were ran. Two high level core surveys, and one thematic survey, which looked more in depth into complicated issues in measuring the progress of digital heritage in Europe.

Looking back on the process, the results of the Thematic Survey are unsatisfactory.

It took a tremendous amount of effort to run different small scale surveys to understand more in depth certain issues at a smaller sample of institutions, but in the end the results were not very helpful to finalise the *ENUMERATE Conceptual Framework*. The lesson learned is that for future surveys in the vein of *ENUMERATE* it suffices to limit the survey to high level core issues. More in depth questions could better be addressed by specialised projects with a clearly targeted sample of participants.

2.6 Many institutions lack management data about digital heritage

As already noticed during the *NUMERIC* project, some types of data are not sufficiently available in the targeted institutions. This became apparent, for instance, where questions were asked about the cost of and budget for digital collections.

After having experimented with more detailed questions in the *Thematic Survey on Cost*, the section covering the cost of digital collections had to be very generic, asking only for estimates rather than precise figures, as many institutions are not able to provide precise information on digitisation budgets. In some cases the *ENUMERATE* Team received notifications that it would actually be possible to provide the information asked for, but that the effort to collect the information at various departments *and* to convert it into the proposed standardised way of delivering it, was considered too big.

The lesson learned is to keep the ambition for high level statistical analysis modest and build in margins to deal with the idiosyncrasies as they exist across the domains. Also, more missionary work needs to be done to improve the availability of management data about digital heritage at the institutional level.

2.7 Finding the right person to address is difficult and time consuming

During the three cycles of the *ENUMERATE* project the fact that memory institutions exist in many varieties affected the survey approach in many ways. The *ENUMERATE* initiative was cross domain, trans-national, and targeted institutions that range from very small (one volunteer; no budget) to very large (hundreds of staff; budgets exceeding €10 million a year). Ideally, questionnaires are sent directly to the staff member that is best able to answer (most of) the questions, but judging from the responses we received in the surveys, there is not one specific profile of a staff member that can be addressed.

The answers were provided by a wide variety of staff members, from policy advisors to collection curators. The less preferable but easiest solution that *ENUMERATE* applied was to address the director in person, with the request to forward the questionnaire internally to the most appropriate staff member. But in many cases the name and contact information of the director were not directly available, so invitations had to be sent anonymously to the secretariats of the directors instead, which decreased the chance of receiving timely responses.

2.8 Contact databases are rapidly becoming unreliable

With the help of the National Coordinators (Gold and Silver scenarios only) an address database was compiled to manage contact information of target institutions. Information gathered on earlier occasions, e.g. the *NUMERIC* project was also re-used. This could contribute to more reliable comparisons over the years, if the same institutions were to participate again. The results of attempts to re-use such contact data were disappointing. For example: within a time span of only 4 years, 25% of the e-mail addresses gathered in the Dutch *NUMERIC* survey from 2008 turned out to be useless.

Similar patterns were found when the *ENUMERATE* Team tried to re-use the contacts from *Core Survey 1* (CS1) in *Core Survey 2* (CS2). Re-use of contact databases turns out to be inefficient unless substantial time and means are invested to keep the contact database up to date. It is more promising to stipulate just-in-time procedures for collecting contacts, irrespective of whether they cover all memory institutions in, for example, a member state or they are selected through random sampling.

2.9 Flexible deadlines increase levels of participation

All three *ENUMERATE* surveys had extended deadlines. This was partly strategy, partly on the request of respondents. The strategy was to have an initial timeframe of a few weeks to respond, which would be accompanied by targeted communication actions. A margin was foreseen at the start of the survey. However in practice the margins had to be extended to reach levels of participation that were needed.

It was the ambition of *ENUMERATE* (as documented in the *Description of Work*), to surpass at least the response rate of the *NUMERIC* survey. This was achieved with both *Core Surveys* (*NUMERIC*: 788 respondents, *CS1*: 1951 respondents; *CS2*: 1375 respondents). During the surveys, many potential respondents indicated that they wanted to contribute, but that they needed more time. During the surveys the response rates were monitored week by week, and the extended deadlines were well used. So flexibility in planning activities is an essential ingredient for running these kinds of large scale surveys.

2.10 Reporting on subsets of data deserve more attention

The Netherlands was the only country to produce a national report based on the *ENUMERATE* data from *Core Survey 1*. This report was produced by DEN, one of the core partners in *ENUMERATE*. It took quite some effort to compile the report, which could look at the data in more detail than was done in the overall *ENUMERATE* report. However, it was worth the effort, the country report was well received in the Netherlands, and its content was re-used in various ways, e.g. in the official report of the Dutch government to the MSEG (as a result, this report was explicitly praised by the European Commission), in the 2013 publication with the overall *Culture Index* of the Netherlands and in several other reports in the digital heritage domain.

There were no spin-off reports produced with analyses about a specific domain (e.g. museums or archives). This is a shame, as the country report of the Netherlands proved that more knowledge can be obtained from a closer analysis than can be done in the overall report on the European level.

2.11 Cross-domain harmonisation deserves more attention

During the project, some meetings were organised to help prepare the *Thematic Survey*. These meetings turned out to be really interesting and useful, and among the most enjoyable parts of the *ENUMERATE* project. Digital heritage statistics is a relatively new subject, and many issues had to be discussed during these meetings on a conceptual level and in an exploratory manner. This is mainly due to the fact that the meetings brought together representatives from the various cultural heritage domains, which have different perspectives and traditions of measuring.

The participants all enjoyed this wider perspective, but also acknowledge that there is more work to be done to bridge the gaps in professional practices across the different domains. It is highly commendable to arrange more of such international cross-domain expert meetings, to strengthen measuring practices in the professional network.

2.12 Measuring digital preservation has not yet reached a mature level

The most complicated theme that *ENUMERATE* addressed was digital preservation. The expert meetings made it clear that the cultural heritage domain is not ready yet to collect statistical data on this topic on a large scale. There are simply not enough operational digital preservation practices. The

lesson learned is that this topic needs closer and more dedicated attention in the future, as *ENUMERATE* has not been able to create a reliable evidence base yet.

2.13 Conclusion

The overall conclusion that can be drawn from the lessons learned, is that a more flexible approach to surveying is desirable, especially if there will not be any centralised resources on the European level anymore in the future. This flexibility is needed to cope with different levels of support in the various countries, with the big differences among the target group (the cultural heritage institutions) and with the extra efforts needed in institutions to acquire the relevant data. In the next chapter the *Conceptual Framework* is presented. It has been set up for future reference and re-use, and takes into account the need for more flexibility.

3 Framework (methodology)

3.1 A definition of the cultural heritage domain in the context of this Framework

The *ENUMERATE Framework* applies only to the European memory institutions (museums, libraries, archives/records offices, audio-visual and film institutes, institutes with curatorial care for archaeology, monuments and sites, and some other/hybrid types of organisations).

The criterion is here that **curatorial care** for (part of) the collections of the institution are **included in its mission**. Institutions that do not hold heritage collections or that have collections of heritage materials (for example books, films, music, etc.) to be lent or sold to contemporary users without the explicit task of safeguarding the collections for future generations, are not included. This leaves out, for instance, school libraries and public libraries without cultural heritage collections.²

A classification of the heritage domain into heritage institution types can be found in **Section 3.7**.

The European dimension originates from the fact that both the earlier *NUMERIC* project and the *ENUMERATE* Thematic Network were funded by the European Commission. However, the *ENUMERATE* surveys were not limited to the 27 countries of the European Union. The EFTA countries were invited to participate as well and through the open survey institutions from other non-EU or non-EFTA-countries participated as well. The full list of countries, as used in *Core Survey 2* questionnaire, can be found in **Appendix 4.4**.

3.2 The size of the European cultural heritage domain

Universe

In the *ENUMERATE* methodology the 'universe' is defined as all memory institutions in Europe. Every institution that belongs to this domain, whether publicly or privately funded, whether actively involved with digitisation or not, should be allowed to contribute to future surveys.

To determine the size of its potential respondent universe *ENUMERATE* had to rely on baseline data from the *NUMERIC* project, since statistical information about the cultural heritage domain on international scale was scattered.³ *NUMERIC* constructed a so-called 'foundation database', summarizing data from various sources, in order to have baseline estimates of the size and composition of the cultural heritage field in the EU member states.⁴

² Where the *NUMERIC* project took great pains to pre-coordinate / to separate institutions in the heritage domain by introducing the problematic concept of the 'relevant institutions', the approach chosen by *ENUMERATE* was to offer some degree of freedom to the caretakers in the field, the National Coordinators, in selecting institutions. The selection of institutions was post-coordinated: in the survey questionnaire one specific question was introduced to draw the line: "Does your institution have collections that need to be preserved for future generations?"

³ EUROSTAT does not yet initiate cross-national research in cultural heritage statistics by itself. The data that are available through Eurostat lack the level of detail needed here.

⁴ Cf. the archived *NUMERIC* website on: <http://www.numeric.ws/> (accessed 14/02/2014).

The National Coordinators were asked to check and correct these initial estimates. Both the estimates from the foundation database and the numbers provided by the national coordinators in the various EU member states are relevant for the procedure described below.⁵

NUMERIC concluded that the universe consists of about 110,000 institutions. For *Core Survey 1*, *ENUMERATE* used this estimate as a starting point. However, this high volume was reduced by restricting the survey to institutions with a mission for collection care and long term curation. The sum total of European memory institutions - the population for the *ENUMERATE Core Surveys* - was estimated to be between 40,000 and 45,000.

The estimates of the *NUMERIC* foundation database and national coordinators for the different cultural heritage institution types from 2008 are given in rows 1 and 2 of **Table 1** (below). Row 3 contains recent (2012) *ENUMERATE* estimates, based on averaging the earlier estimates while skipping the top outliers for all classes but the museums, where outliers were less apparent. The estimate for the class of 'Other institutions' is a set number. Public libraries are left out.

In order to arrive at a realistic target number for surveys under this framework, the decision was taken to reduce the various types of heritage collections to the parent/umbrella classes of: Libraries, Archives, AV institutes, Museums and Others (the latter including Monuments and Sites).

Table 1: The size of the cultural heritage domain

	Archives and Record Offices	A-V / Film Institutes	Libraries (all)	Museums (all)	Other	All Institutions
Foundation estimate	766	29	30328	14299	-	45422
National coordinator estimate	5194	257	23568	14853	77	43949
<i>ENUMERATE</i> estimate	2023	93	25413	14577	875	42981

As far as we know, the situation has not changed much since 2012, so for the present situation (January 2014) the figures are left unaltered. It is however recommended to organise a regular check of these data. This could be done by asking national representatives to make updated estimates of the national heritage domain, for instance in the two-yearly *Progress Reports to the Member States Expert Group of the European Commission*.

Institutions taking part in the surveys will be able to classify themselves according to a more detailed list of institution types (see **Section 3.7**). As a consequence of this, data collected may be statistically valid for these broad types of institutions in the EU as a whole, provided that an accepted sampling routine is chosen and that survey targets (response rate) will be such that the calculated sample sizes are met.

⁵ In preparation of the *Core Survey* the National Coordinators will again be asked to make new estimates of the number of heritage institutions in their countries. The data thus collected may alter some of the numbers, but will not essentially change the methodology proposed here.

Survey targets

This conceptual framework essentially covers **all** European memory institutions as defined above. Any actual survey effort, either for the whole European domain or any subsection of it, should ideally make use of a contact database containing the necessary details for *all* memory institutions in its specific sub-domain. If that can be achieved no sampling methodology is necessary. Representativeness can be calculated afterwards. This is the state of affairs as covered by the Gold scenario as used for *ENUMERATE Core Survey 2* (see **Section 3.11**).

In *ENUMERATE Deliverable D2.4*, the *Core Survey 1 methodology*, a comprehensive procedure was introduced to calculate targets per participating country. The starting point was to have representative samples per high-level class of heritage institutions *for the EU as a whole*. Based on that purpose, targets for the *ENUMERATE Core Surveys* were calculated [refer to D2.4, pp13-15]. **Table 2** presents samples per high level cultural heritage domain and across the EU.

Please note that this sample size refers to the required response rate in a random sampling set-up, not to the amount of institutions that can be invited.

Table 2: The samples per cultural heritage domain and across the EU

	Archives and Record Offices	A-V / Film Institutes	Libraries (all)	Museums (all)	Other	All Institutions
<i>ENUMERATE</i> estimate	2023	93	25413	14577	875	42981
Sample size	323	76	379	375	268	1421

In the bottom row are calculated samples for the corresponding institution types. These sample sizes are calculated according to standard statistical procedures. It was decided that a confidence interval/error margin of 5% would be acceptable. The confidence level was set to 95%. With an estimated size of the total EU population of 2023 (as in the case of archives and record offices) the sample can be calculated using one of the online available sample size calculators as 323.⁶

The same procedure is followed for the other cultural heritage domains. Although there is uncertainty about the exactness of these estimated population data, the sum total (1421) can serve as a starting point to calculate the size of target samples for each EU member state. Depending on the actual response, which may be higher than the targets set here, statistically valid results for subtypes of institutions (e.g. museums of art, museums of technology, etc.) on the EU level should be possible.

If the response is lower than the target, statistically valid results may only be possible on cumulative levels (generic institution types, country level or maybe even just the European level). The reliability of the data gathered needs to be addressed in the data analysis phase.

Both in *Core Survey 1* and *Core Survey 2* the national coordinators were asked to strive for specific survey targets. These targets were fixed partly based on statistical principles, getting representative data for the European cultural heritage domain as a *whole*, partly on pragmatic grounds (the work had to be done by volunteers).⁷ In the table below the distribution of these targets is made clear.

⁶ See for instance <http://www.allesovermarktonderzoek.nl/steekproef-algemeen/steekproef-berekenen> (accessed 14/02/2014) or <http://www.surveysystem.com/sscalc.htm> (accessed 14/02/2014).

⁷ The rationale behind these figures is explained in deliverable D2.4 *ENUMERATE Core Survey 1 Methodology*.

Table 3: The sample of Archives / Record Offices distributed over the EU countries according to the percentage of total EU heritage domain per country

Country	ENUMERATE sample	Archives/ Record Offices	Audio- Visual / Film Institutes	Libraries	Museums	Other
Austria	24	6	1	7	6	5
Belgium	38	9	2	10	10	7
Bulgaria	26	6	1	7	7	5
Cyprus	15	3	1	4	4	3
Czech Republic	53	12	3	14	14	10
Denmark	35	8	2	9	9	7
Estonia	15	3	1	4	4	3
Finland	33	7	2	9	9	6
France	127	29	7	34	34	24
Germany	150	34	8	40	40	28
Greece	35	8	2	9	9	7
Hungary	44	10	2	12	12	8
Ireland	15	3	1	4	4	3
Italy	150	34	8	40	40	28
Latvia	15	3	1	4	4	3
Lithuania	22	5	1	6	6	4
Luxembourg	15	3	1	4	4	3
Malta	15	3	1	4	4	3
Netherlands	67	15	4	18	18	13
Poland	125	28	7	33	33	24
Portugal	27	6	1	7	7	5
Romania	46	10	2	12	12	9
Slovakia	25	6	1	7	7	5
Slovenia	15	3	1	4	4	3
Spain	103	23	5	27	27	19
Sweden	38	9	2	10	10	7
United Kingdom	150	34	8	40	40	28
	1421	323	76	379	375	268

In the second column are the target totals for the 27 EU member states. The minimum target value is set to 15. The maximum target value is set to 150. The columns on the left are **indicative** of the

distribution of these target values over the separate heritage domains. Small deviations in the numbers are due to the rounding of numbers and the set minimum value of 1 for each individual cell in the table.

Where in the future modular survey efforts - in particular: per EU member state - will be most probable, it is advisable to re-calculate targets along the procedure described above for each and every sub-domain (EU member state/country). This would justify the effort of organising surveys by making outcomes useful at the level of the initiator.

3.3 Baseline data collected in two subsequent Core Surveys

The *ENUMERATE Core Survey* was first organised early in 2012, and was repeated during the autumn of 2013. Since the *ENUMERATE* initiative aimed at both methodology development and the collection of baseline data most of the topics covered and questions asked were, if deemed necessary, only slightly changed in the second iteration. Most questions remained unaltered in essence during both data collection efforts.

The survey processes have generated a very significant response, representing thousands of cultural heritage organisations across Europe, and providing both baseline and trend data alongside in-depth qualitative information.

Survey data were cleaned and harmonised, and were made available in various formats (e.g. SPSS, and Excel).

Collectively, this data has been assembled into a body of knowledge, analysis and raw-data that is available through the *ENUMERATE Data Platform*. This *Data Platform* was constructed by consortium member Digibis as a core deliverable of the funded phase of work. It has been developed with specific functionality to support the re-use of the *ENUMERATE* data. The data platform itself represents a significant investment of intellectual capital, having been developed to support specific use cases arising from the reporting needs of different communities at institutional, regional, national and EU-wide levels.

3.4 Target audience and reporting formats

The main audiences for the data on digital heritage collected in the *ENUMERATE* project are considered to be:

- EU (European Commission; Europeana);
- National governments in EU member states and affiliated countries;
- Management of cultural heritage Institutions.

In addition to these primary audiences the results of the *ENUMERATE* core surveys will be of interest to:

- Individual professionals working in the cultural heritage domain;
- Commercial parties developing products for the field of digital cultural heritage;
- Academic researchers with a scientific interest in the accumulated data;
- International organisations.

In the *ENUMERATE Description of Work* it was included that after each survey a *Survey Report* presenting aggregated outcomes would be published, while the raw survey data would be made accessible through the *ENUMERATE Data Platform*. Unlike *NUMERIC* it was decided to separate the documentation about the methodology completely from the results of the surveys.

In the *Core Survey 1* report a modest comparison could be made with the data collected in the *NUMERIC* project, after *Core Survey 2* more extensive trend analyses were possible. It is obvious from the diversity of target audiences, that publicly available reports and presentations need to be widely announced through the various communication channels, both specific like media directly linked to the project and generic (social) media..

Apart from these general survey reports, a feedback form was developed to give back the answers given by individual institutions, for future reference. This was also meant as a 'thank you' to the institution for participating in the survey. Also a generic and anonymous benchmarking tool was developed as part of the *ENUMERATE Data Platform*.

In future monitoring projects it is advisable to develop multiple reporting formats that are closely aligned with the analysis of the target audiences. We suggest the following reporting formats:

Audience	Interest	Reporting formats
European Commission	Data are needed for supporting policy decisions in the field of (digital) cultural heritage in the EU.	<i>Survey Report on Digitisation in European Cultural Heritage Institutions</i> ; Raw data on <i>Data Platform</i> ; Dedicated analyses based on the raw data (on demand)
Europeana and other Supranational / Cross domain (aggregating) initiatives	Data are needed to trace gaps in the collecting of digital content and/or signalling potential resources	<i>Survey Report</i> ; Trend reports based on raw survey data; Information about the collections of possible content providers* [this is of course problematic due to the promised anonymity of data]
National Ministries and Agencies	Data are needed for supporting policy decisions in the field of (digital) cultural heritage on a national scale.	National reports on digital heritage collections; Reports offering context to national performance data
Cultural Heritage Institutions	Data are needed to develop institutional policies and put internal indicators of progress into context.	Benchmarking reports (e.g. on a domain level or on the level of a specific type of institution, like art museums) offering insight into specific indicators; Dashboard-like presentations.

3.5 Minimum set of (core) indicators for 'high level'-surveys

ENUMERATE focused on high level data that could be used to understand the overall progress in the area of digital heritage. Presented here are high-level indicators of critical factors in the development and management of digital heritage collections. These indicators enable collecting headline statistics. They will probably be valid during a limited period of time (3-5 years), but continuous monitoring of the development of digital collections is necessary. Procedures for the assimilation of new indicators are presented in **Section 3.8**.

An additional remark must be made about the way these indicators are formulated. We have taken care to describe the indicators from the perspective of the individual institutions and **not** from the perspective of the higher-level stakeholders (e.g. National Governments, European Commission). So the indicator is "Presence of digital collections" and not "Proportion of institutions with digital collections". Both descriptions are consistent.

The indicators are ordered along the major themes of the *ENUMERATE Description of Work* (pp10-11):

- Growth of digital collections (supply);
- Usage of digital material (demand);
- Costs of digitisation (economics);
- Digital preservation practices (sustainability)

First the minimum set will be named, which is actually an outcome of the various activities during the full three year period of the *ENUMERATE Thematic Network*, followed by some more in-depth indicators. The minimum set can be translated straight away into the questionnaire that was in use in the last survey round (*Core Survey 2*).

Where the denominator ‘digital collections’ is used below, it is always about ‘digital collections of heritage materials’. For quick reference we have added between square brackets the number of the *Core Survey 2* question [q.n] expressing the indicator.

Indicators:

Growth of digital collections (supply);

1. Presence of digital heritage collections [q.12];
2. Presence of born digital collections [q.14];
3. Existence of embedded policies related to expanding digital collections [q.13];
4. Degree of descriptive metadata cataloguing in a digital collection database [q.16];
5. Degree of digital reproduction/representation presence in the digital collections [q.17];
6. Necessity to reproduce analogue heritage collections in digital format (per object type) [q.18].

Usage of digital material (demand):

7. Presence of institutional website(s) [q.4 & q.23];
8. Importance of different types of use of digital collections [q.19];
9. Existence of embedded policies related to the use of digital collections [q.20];
10. Status of monitoring the access to and use of digital collections [q.21];
11. Use of monitoring methods [q.22];
12. Popularity of network access options regarding digital collections [q.23].

Costs of digitisation (economics):

13. Annual expenditures on digital collections [q.26];
14. Number of staff engaged in creating and preserving digital collections [q.29;]
15. Sources of funding for digital collections [q.31];
16. Ratio of incidental versus structural costs of digital collections [q.27];
17. Ratio of in-house versus outsourced costs of digital collections [q.28];
18. Relative weight of costs associated with creating and preserving digital collections [q.28+].

Digital preservation practices (sustainability):

19. Existence of embedded of policies related to the sustainability of digital collections [q.24];
20. Adherence to international standards in digital preservation practice [q.25];

The indicators are high level pointers to the state of the art in the various memory institutions. A more in depth description of the indicators (questions, disaggregation notes, data sources and statistical notes can be found in **Appendix 4.1**.

3.6 More in-depth indicators for thematic surveys

During the lifetime of the *ENUMERATE* Thematic Network several more in-depth indicators were found in the literature, or have been formulated. These were discussed at several occasions, for instance during the four *ENUMERATE* Expert meetings. Where deemed useful a selection of these indicators were translated into the questionnaires developed for the *Thematic Survey* phase. One of the lessons learned was that these more in-depth indicators do not really fit well in large-scale surveys set out across all countries in various languages in all domains. They are presented here to be used in more modular, specialised or targeted types of surveys.

A few examples of more in-depth indicators:

- Growth of digital collections (supply):
 1. Number of metadata records in Collection Registration System⁸ (CRS) (per object type);
 2. Physical storage needed for digital collections (also: per object type);
 3. Number of born digital units in entire digital collection (also: per object type).

- Usage of digital material (demand):
 4. Online availability (y/n) of the CRS;
 5. Online availability of the contents of CRS (per object type);
 6. Adherence to *Conditions of Use*.

- Costs of digitisation (economics):
 7. Specification of any cost item related to an institution's digital collections in the annual report;
 8. Specification of **incidental** cost items (per size of digital collection);
 9. Specification of **structural** cost items (per digital collection).

A database of potential survey questions has been built up and will be transferred to the group or organisation that will be responsible for follow-up of the *ENUMERATE* Thematic Network.

3.7 Statistical standards applying to *all* indicators

In **Section 3.1** the *ENUMERATE Framework* was described as covering the European memory institutions (museums, libraries, archives/records offices, audio-visual and film institutes, institutes with curatorial care for archaeology, monuments and sites, and some other/hybrid types of organisations). Below some issues about the breakdown of the full domain are discussed.

Unit level issue

The 'memory institution' is thus the principal entity we shall consider here. Definitions of this class/entity and its subclasses (e.g. museums and libraries) are in the main definition list. A thorny problem with institutions of this kind is that they can be complex organisations with often a hierarchical administrative structure that may obscure at what level indicators should be collected. The institution may be housed in one building, or it may be an aggregate of several physical sites under the umbrella of one administrative unit. Consider for example the question concerning an institution's total annual budget.

⁸ Also called the Collection Management System (CMS).

It is suspected that certain higher education libraries may have filled in the total annual budget for the entire library or even the entire university or polytechnic, whereas the intended amount should have been restricted to the budget of the cultural heritage related unit.

It is proposed that the concept of 'service point' as the standard statistical unit should be used, under the condition that it should be recognisable as a named entity.⁹ Only where this is explicitly stated in the description of the indicator we recommend the use of the administrative unit.

Part of the complications arising from the fact that institutions can be very dissimilar regarding their hierarchical organisation will be overcome by classifying institutions in a few size categories. This way it will be possible to do analyses on heritage institutions in the same size category.

Classifications

1. Type of Institution

A more elaborate classification of the 'memory institutions' covered by this framework into types is as follows:

- National archive;
- Other archive / records office;
- Audio-visual / broadcasting archive;
- Film institute;
- Institution for performing arts;
- Museum of art;
- Museum of archaeology or history;
- Museum of natural history or natural science;
- Museum of science or technology;
- Museum of ethnography or anthropology;
- National library;
- Higher education library;
- Public library;
- Special or other type of library;
- Institution for monument care;
- Hybrid type of institution, serving several cultural heritage domains
- Other (to be specified).

In the last core survey (*Core Survey 2*) a rather high percentage (19%) of respondents opted for the class of 'Other'. Since most institutions are actually hybrid institutions we propose to leave out this category 'Other' in future surveys. For practical reasons in both the *NUMERIC* survey and the *ENUMERATE* suit of surveys a rough classification into a few broader classes was used in the *analysis* of survey results. This standard high level classification is as follows:

- Archives/record offices;
- Audio-visual, broadcasting or film institutes;
- Libraries;
- Museums;
- Other types.

2. Size of Institution:

In the *ENUMERATE* Framework the size of an institution is determined by the total annual budget for that organisation as published in its last annual account. This total annual budget may include government funding, project funding, revenues from commercial activities, and other sources.

⁹ See: ITU, *Measuring the WSIS Targets A statistical framework* (2011), p. 108-109.

Another indicator of size that is frequently used relates to the number of paid staff working in the institution. Because the mix of paid staff and volunteers may be confusing, for this indicator, though actually being canvassed in the *Core Survey*, no further classification is made.

In two subsequent iterations of the *Core Survey ENUMERATE* Team used:

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

Please note that the unit level issue is decisive here. We found for example that certain survey outcomes in the *Core Survey* were ambiguous: despite clear instructions some higher education libraries apparently filled in the annual budget for the entire university. Cf. “If your institution is part of a larger organisation (e.g. a higher education library that is part of a higher education institution) only provide the budget of the cultural heritage related unit.”

3. Type of object

Depending on the response in practice classes that are poorly represented in survey outcomes can be classed under ‘Other types’.

We recognise that the classifications stipulated above link back to the reality of how collections are composed of specific types of heritage materials. Part of the *ENUMERATE* research was to compile a hierarchical list of heritage materials. Because the list is rather lengthy, it is moved to **Appendix 4.2**.

Disaggregation

In line with the classifications stipulated above, survey results can be disaggregated according to these general classes of institutions:

- Type / Domain of institution / organisation: museum, library, etc.
- Total annual budget of institution / organisation: <€10,000; €10,000-50,000, etc.
- For international analysis/comparison purposes: Country where the institution is located.

3.8 A procedure for the assimilation of new indicators

During the *ENUMERATE* project, various parties made enquiries whether *ENUMERATE* could provide statistical data on a specific topic, or they provided suggestions for topics to be included in the *ENUMERATE* surveys. Those suggestions included topics like copyright (e.g. total number of orphan works in Europe) and general use of ICT (how many museums provide free wifi?).

To provide continuity after the *NUMERIC* project, *ENUMERATE* was quite clearly scoped with the four main topics as described (size, cost and use of digital collections and digital preservation). However, it will be interesting to understand existing needs for reliable statistical information about digital heritage that is not provided by *ENUMERATE*. Some research into these needs is highly recommended.

The decision to include new indicators and questions in this *Conceptual Framework* should be based on the following criteria:

- The new indicators are relevant for the whole population of the cultural heritage domain as presented in this framework (**Section 3.1**);
- The statistical standards that apply to all accepted indicators also apply to the newly suggested indicators (**Section 3.7**).

3.9 Definitions and Tools

In the course of *ENUMERATE*'s three-year survey project an extensive list of basic definitions has been developed. In the different Core and Thematic Surveys that were run selections from the main list were made. The full list of English definitions is included here for direct reference in **Appendix 4.3**. The definition list in some other European languages can be found on the *ENUMERATE* website (section Guidance).

On the *ENUMERATE website and Data Platform* several other tools have been published during the project lifetime to support the process of running a Europe-wide survey. The tools may be useful for the organisers of a new survey, for national coordinators, for potential respondents and for re-users of the raw data that *ENUMERATE* has collected and published. Because some of these tools are extensive documents they have not been included as appendices in this framework. They are available online on the *ENUMERATE* website, www.enumerate.eu. The tools are:

- The *ENUMERATE* questionnaires in linear format (as PDF files);
- An overview of harmonisation tools, which provides an overview of validation and harmonisation tools concerning terminology, digitisation costs, collection types and web-statistics;
- An overview of national and international monitoring initiatives in the area of digital heritage;
- An inventory of software tools for the analysis of the raw survey data;
- A charter of the National Coordinators;
- An overview of the National Coordinators that helped to publish the surveys across Europe;
- Examples of letters to be used as invitations or incentives to potential respondents.

In 2014/2015, these tools will be migrated from the *ENUMERATE* website to the Europeana Pro platform (pro.europeana.eu), as part of the Europeana v3 project. Where useful or possible, the tools will also be updated then.

3.10 Liaising with other monitoring practices

During the project lifetime, *ENUMERATE* has been in touch with several other initiatives to collect statistical data on cultural heritage and related domains (e.g. *ESSnet*, *EGMUS*, *WSIS+*, *Prestocentre*, *OpenCultureData*, and TEL). It became clear that in these other initiatives methodology development was also still very much part of the work done. This made it difficult to make clear arrangements about, for instance, sharing and re-using each other's data. It became clear that the type of work that *ENUMERATE* tried to achieve is also still very much in development in related areas, and that future action is needed to make new attempts for aligning the different results. See also the overview of national and international monitoring initiatives as mentioned in the previous paragraph for possible initiatives to liaise with in future survey activities.

3.11 Suggestions for the organisation of future data collection activities

At the end of paragraph 2 of **Lessons learned** it was concluded that a more flexible approach to surveying is desirable, especially if in future years there will not be any centralised resources for monitoring developments in digital cultural heritage at the level of the EU.

From the perspective of the *ENUMERATE* consortium, the ideal situation would be to include in future data collection efforts **all** institutions conforming to the definitions as set in **Section 3.1**. Digitisation of cultural heritage is only in a limited way domain specific.

A cross-domain monitoring approach is preferable. This means that there continues to be a need for articulating the full European population at the level of the individual countries. If the cross-domain approach is not feasible, a more limited scope of one or more memory institution types within an individual country could be an alternative. In the *ENUMERATE* surveys we learned that pragmatic compromises were sometimes necessary because the project made use of a network of volunteers (i.e. the National Coordinators), representing their countries, either on a government level or from the cultural heritage sector. Their personal network had a strong influence on way potential respondents could be reached.

As was stated above the *ENUMERATE* Team experienced that realising substantial coverage cannot easily be done in a standardised way if substantial means to support the operation are lacking. Therefore we must get into more detail about possible scenarios for running surveys in different countries and the sampling methods that are needed to compromise for the lack of complete contact databases.

Three scenarios

In *Core Survey 1* it turned out that to identify possible respondents, memory institutions, was an easy job in a few countries and a hard one in others. This made it difficult to find a one-size-fits-all sampling approach. Therefore it was decided to use in *Core Survey 2* three levels of involvement for national coordinators, three sampling scenarios: *Gold*, *Silver*, and *Bronze*.

The scenarios differ in the way the population is covered:

- **Gold:** Contact details for the whole population of all types of memory institutions are available. All institutions are invited to participate.
- **Silver:** At best contact details for memory institutions in only some types of institutions (e.g. archives) are available. For the other types maximally representative samples must be constructed (see below), e.g. by procedures like selecting every n^{th} organisation on a list of memory institutions or making the best of *ad hoc* collecting contact details. In this scenario guidance to the National Coordinators in constructing lists of contact details, was needed.
- **Bronze:** A fully open call to participate is made on various different platforms. In this scenario a national representative/coordinator is still essential in announcing the survey and collecting the contact details of institutes that sign up. Here too the *ENUMERATE* team can offer assistance.

In all scenarios some form of central administration and guidance at the European level is desirable.

Whatever the scenario chosen in a particular country, future monitoring projects can reuse the available contact details of respondents in earlier EU funded surveys (*NUMERIC*, *Core Survey 1*, *Core Survey 2*, and *Thematic Survey*), bearing in mind that the lesson learned is that within a few years, a large portion of the contact database will become useless if not maintained (see par. 2.8).

In the *Gold* scenario the respondents to previous surveys will be addressed by default. In the *Silver* and *Bronze scenarios* the successor to *ENUMERATE* will compile a basic file and the national coordinators will be asked to check whether these addresses are still valid.

Opting for either the *Gold*, *Silver* or *Bronze scenario* is not necessarily related to realising representativeness of survey results per participating country. Although it is highly unlikely, one might invite all memory institutions in a particular country to do the survey and yet get zero response. This is where the role of the national contacts becomes decisive. That being said we may cautiously state that the benefits of opting for *Gold* are potentially more advantageous than opting for one of the other scenarios:

- In a country opting for the *Gold scenario* it will be easier to draw statistically valid conclusions at the level of the broad heritage domains: Archives, Audio-visual institutes, Film institutes, Libraries, Museums, and Institutes for Archaeology and Monument Care. In the *Gold scenario*, for example, it may be possible to conclude that monitoring the actual use of digital collections for a specific country is more common in e.g. Museums than it is in Film institutes.
- In countries where the *Bronze scenario* is chosen, data collected can contribute to the overall picture of digital heritage collections in the EU, but it will be improbable that one can do reliable statements about the national state of affairs.

Sampling (*Silver scenario*)

In as far as sampling is necessary, more in particular in the *Silver scenario*, the aim should be to have a selection of institutions which broadly reflects the cultural heritage domain in the various European countries. To ascertain that the maximum number of potential respondents is reached the national contacts must be asked to draw up lists of institutions for the main heritage domains in their country.

It is important that these lists cover as much of the heritage domain as possible, both in terms of geographical distribution and size. Proportional numbers of large, medium-sized and small institutions should be included. Furthermore, among the selected institutions both digitising and non-digitising institutions should be represented. The *ENUMERATE* Framework is not about the *digitising* institutions alone!

Here are some examples of compiling lists of institutions *in the Silver scenario*:

Example 1

A directory of central government and municipal archives is located. In this example, all 100 archives in the directory are listed by regional / county / departmental responsibility. After investigation, you estimate that half of the list could be approached to take part in the survey. The procedure could be to sort all archives alphabetically and then count down this list, selecting every 2nd institution. Or, there are, for example, 28 regions in your country. You incrementally select regions - constantly ensuring a good geographic spread - and include all listed archives from each region until a total of 50 have been reached.

Example 2

There does not appear to be a conveniently available list of all museums nationwide. However, a list does exist of museums that participated in an earlier survey on collection mobility. This lists 60 institutions and there is no breakdown of the type of such museums. You decide to address all institutions on the list.

Example 3

No central list of higher education libraries exists, but you experience that it is fairly easy to trace libraries in the universities and graduate schools of your country. Since the number of these institutions would amount to 150, you decide that covering 1/3 would be feasible. 50 institutions are selected by examining a map to ensure a good geographic spread across the country.

In any coordinated survey lists of institutions to be invited will have to be compiled by the coordinating party. These lists should cover the following details (as column headings):

Institution type; Name of contact person; E-mail address

If access to personalised address lists is problematic, we advise to address the management (e.g. director or CEO) of the cultural heritage institutions.

3.12 Planning for future surveys

The time period separating the two *ENUMERATE Core Surveys* was about 18 months. The ideal period covered by data collection efforts would be one year. This is in line with most internationally agreed practices. This framework proposes to relate any question about activities over an extended time span to the period, usually the year, covered by the last annual report.

For the time being, for pragmatic reasons, this *Framework* proposes to opt for a two-year interval of analyses. The outcomes of any partial survey effort should relate to this timetable.

Additional information and recommendations about continuation of the *ENUMERATE Framework* through new surveys can be found in the *ENUMERATE Deliverable D4.8: Business Plan and Sustainability Strategy*.

4 RECOMMENDATIONS

The monitoring of the growth, cost, use and sustainability of digital collections is an intricate process of relating the local interests of the heritage institutions to the global perspectives sought by governments and overarching associations and organisations. Recommendations should be targeted at the appropriate audience.

Below the main recommendations from the ENUMERATE initiative are listed. For the sake of clarity two classes of stakeholders will be considered:

1. Individual cultural heritage organisations (museums, libraries, archives: the 'bottom-up' perspective)
2. Governments and umbrella organisations (governments, associations and the like)

4.1 Recommendations for Cultural Heritage Institutions

One of the Lessons Learned was that many institutions lack management data about their digital collections (§2.6). Still knowing how well your organisation is doing in fulfilling its mission in the digital realm is key to being accountable in the long run.

A centrally organised survey is probably not the most appropriate incentive to start collecting the kind of data necessary to assess the performance of your institution. Therefore we recommend that heritage institutions begin collecting management data in a systematic way and at regular intervals as soon as the urgency of doing so is recognised and widely accepted within the organisation.

Below is presented a list of recommendations that may help individual institutions in achieving accountability in this respect.

4.1.1 ***Take part in trusted surveys, organised with the aim of monitoring the size, growth, cost, use and sustainability of digital heritage collections***

It goes without saying that taking part in surveys such as the ENUMERATE Core Surveys is helpful both for the community of museums, libraries and archives, and or the individual heritage institutions. The ENUMERATE Team repeatedly received notes affirming the usefulness of over-viewing the status of an institution's digital collections.

4.1.2 ***Decide at management level about whether or not performance indicators related to the digital heritage collections of your institution should be collected***

Because collecting these data must be planned and facilitated over a longer period of time, commitment of the management is key. It is recommended that one of the staff members be made responsible for internal monitoring.

4.1.3 ***Make an inventory of (a limited number of) performance indicators that are or could be important for your institution***

It may help to consider heritage institutions similar to your own and decide about the indicators most suitable for benchmarking.

4.1.4 ***Compare the indicators mentioned above with the indicators proposed in national and/or international context, in particular those of the ENUMERATE Framework***

The relaunch of the ENUMERATE initiative as part of Europeana v3 will probably guarantee that for the time being the list of indicators will be maintained/updated. (§3.5)

4.1.5 ***Select the performance indicators with the highest fit and viability for your organisation***

These are the indicators that allow you to monitor how well you are doing as compared to other institutions in the field. It should be taken into account whether the systematic collection of these

indicators is feasible within your organisation and whether (information) systems are in place to automatically collect the necessary data.

4.1.6 *Make or adapt precise definitions and descriptions of the selected indicators and adjust these to international standards or best practices if possible*

This step includes the precise description of the indicators, but also the preferred frequency of collecting. Conforming to standards or best practices raises opportunities to make comparisons. It could also be your contribution to strengthening the international evidence base in the domain of museums, libraries and archives.

4.1.7 *Describe the way in which the selected performance indicators will be collected within your institution*

In this step it should be outlined how the necessary data are going to be collected; who is responsible for collecting; and what tools are necessary to do the job (e.g. *Google Analytics*).

4.1.8 *Decide on delivery and storage of outcomes*

It is evident that some outcomes of internal monitoring are not suitable for publication. For that reason survey outcomes are only published anonymously and as aggregated results. But whether or not to publish your performance indicators is a different issue. It is in the interest of the community of museums, libraries and archives that institutions do not work in isolation and assimilate and share lessons learned where possible.

4.1.9 *Start up the implementation process*

The activity of collecting data for performance indicators is actually a cyclic process. Once it has been initiated evaluation and modification should be part of the process (plan-do-check-act).

4.2 Recommendations for Governments and Umbrella Organisations

The overall conclusion that can be drawn from organising the ENUMERATE surveys is that a more flexible approach to surveying is desirable, especially if there are no centralised resources available. This flexibility is needed to cope with different levels of support in various geographical regions; with the big differences among the target group (the cultural heritage institutions); and with the extra efforts needed in institutions to acquire the relevant data.

Below is presented a list of recommendations that may help survey organising parties to do their job.

4.2.1 *When organising surveys give priority to a more modular approach*

For future surveys, we would like to recommend a more modular approach, to be more flexible and adaptable in collecting relevant data. However, a central place where the data can be collected, analysed and published will still be needed. *ENUMERATE* sees the *Europeana Pro* platform as a good vehicle to this end for European surveys. (§2.1)

4.2.2 *Consider customisation options to increase the level of participation in survey initiatives*

It is considered crucial to continuously make incentives explicit and invest in methods to reduce the workload for contributors to the survey process as much as possible. (§2.4)

4.2.3 *Keep the ambition for high level statistical analysis modest and build in margins to deal with the idiosyncrasies as they exist across the domains*

More missionary work needs to be done to improve the availability of management data about digital heritage at the institutional level. In some cases the *ENUMERATE* Team received notifications that it

would actually be possible to provide the information asked for, but that the effort to collect the information at various departments *and* to convert it into the proposed standardised way of delivering it, was considered too big. (§2.6)

4.2.4 Use ENUMERATE's estimates of the European cultural heritage domain in future European surveys

The Framework contains a detailed description of the European cultural heritage domain. For analyses on the level of the parent classes of Libraries, Archives and Record Offices, A-V and Film Institutes, Museums and Other the estimates made by ENUMERATE may be used. (§3.2)

4.2.5 It is recommended to regularly check ENUMERATE's estimates of the European cultural heritage domain on country level

Evaluation of the estimates could be done by asking a national representative to make updated estimates of the national heritage domain, for instance in the two-yearly *Progress Reports to the Member States Expert Group of the European Commission*. (§3.2)

4.2.6 Adapt sampling size in accordance with (new) estimates of the heritage domain per country for obtaining reliable results on country level

Where in the future modular survey efforts - in particular: per EU member state - will be most probable, it is advisable to re-calculate targets along the procedure described above for each and every sub-domain (EU member state/country). This would justify the effort of organising surveys by making outcomes useful at the level of the initiator. (§3.2)

4.2.7 Take into account that building up contact databases and finding the right person to target is difficult and time consuming

Ideally, questionnaires are sent directly to the staff member that is best able to answer (most of) the questions, but there is no specific profile of a staff member that can be addressed across institutions. A feasible solution is to address the director in person, with the request to forward the questionnaire internally to the most appropriate staff member. But even compiling a database with management contacts was time consuming. (§2.7 and 2.8)

4.2.8 Use ENUMERATE definitions and tools in future surveys on digital heritage collections

An extensive list of basic definitions has been developed. The full list of English definitions is included here for direct reference. (§3.1, §3.9, Appendix 4.3)

4.2.9 Use ENUMERATE statistical standards in future surveys on digital heritage collections

These standards relate to the statistical unit to consider (the memory institution as a 'service point'), the classifications to be used (the types and sizes of institutions, the types of heritage objects) and proposed disaggregation options. (§3.7)

4.2.10 Implement ENUMERATE (core) indicators in future surveys on digital heritage collections

These indicators enable collecting headline statistics. A preliminary elaboration of these indicators is in the Core Survey questionnaire that was in use in the last ENUMERATE survey round. (§3.5)

4.2.11 Re-evaluate the set of core indicators and modify according to the evolving needs for reliable statistical information about digital heritage in future surveys

Research into these information needs is highly recommended. The Framework is not intended to be a static series of prescriptions. To maintain cross-domain validity make sure that new indicators are relevant for the whole population of the cultural heritage domain as defined here. (§3.1, §3.8)

4.2.12 *Relate future survey efforts to the annual reporting cycle of the cultural heritage institutions*

The ideal period covered by data collection efforts would be one year, which would make it possible to relate any question about digital collection activities to the year covered by the latest annual report. For pragmatic reasons (organisational efforts, perceived survey fatigue) this Framework proposes to opt for a two-year interval of analyses. The outcomes of any (partial) survey effort should relate to this timeframe. (§3.12)

4.2.13 *Take into account that translating questionnaires into multiple languages is a project in itself*

The time the ENUMERATE Team had to spend on planning and producing translations of the Core Survey questionnaire was substantial. The team had to compromise on the quality of the translations. A thorough procedure would provide in quality checks, which might easily result in a doubling of efforts. (§2.1)

4.2.14 *Remember that flexibility in planning activities is an essential ingredient for running large scale surveys*

During the ENUMERATE Core Surveys the response rates were monitored week by week, and the extended deadlines proved to be remunerative. (§2.9)

4.2.15 *In future surveys develop multiple reporting formats that are closely aligned with an analysis of the target audiences and their information needs*

The Conceptual Framework contains an example of how audiences and reporting formats can be described. (§3.4)

4.2.16 *Cross-domain harmonisation deserves more attention*

It is highly commendable to arrange more international cross-domain expert meetings, to strengthen measuring practices in the professional network. (§2.11)

4.2.17 *Initiate new attempts in aligning different monitoring practices*

The ENUMERATE Team learned that the type of work the project tried to achieve is still very much in development in related areas. This complicated the making of clear arrangements about, for instance, sharing and re-using each other's data. (§3.10)

4.2.18 *Foster research in measuring digital preservation*

Measuring digital preservation has not yet reached a mature level The cultural heritage domain is not yet ready to collect statistical data about digital preservation on a large scale. The topic needs closer and more dedicated attention in the future. (§2.12)

5 APPENDICES

5.1 Core indicators for ‘high level’-surveys

For all concepts defined here, see the list of definitions in **Appendix 4.3**.

Growth of digital collections (supply)

1. Presence of digital heritage collections

Explanatory note: The indicator allows institutions to compare “digitisation activity” very broadly with other institutions. On an aggregated scale it provides reference materials for progress in the domain of digital collections (e.g. *Proportion of memory institutions with digital heritage collections*).

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Useful are disaggregation by type of institution; by size of institution.

Remarks: Once the proportion approaches 100% in a specific domain, it will not be a useful indicator anymore.

Example question (Core Survey 2, 12): *Does your organisation have *digital collections* or is it currently involved in collection *digitisation* activities? [Yes/No]*

2. Presence of born digital collections

Explanatory note: The indicator allows institutions to compare performance in collecting any kind of born digital heritage materials (i.e. software, digital documents, digital art, harvested web content, etc.) with others. On an aggregated scale it provides reference materials for progress in the growth of born-digital collections (e.g. *Proportion of memory institutions with born-digital materials*).

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Once the proportion approaches 100% in a specific domain, it will not be a useful indicator anymore.

Example question (Core Survey 2, 13): *Does your organisation collect born digital heritage? [Yes/No/Do not know] Answer this question with ‘yes’ if your institution collects any kind of *born digital heritage* materials (i.e. software, digital documents, digital art, harvested web content, etc.) with the explicit intention of preserving these born digital materials for future generations.]*

3. Existence of embedded policies related to expanding digital collections

Explanatory note: This indicator is a measure of the support for activities related to the growth and maintenance of digital collections within an institution.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Two main difficulties have been encountered and were widely discussed with respect to this indicator. The first is that policies may be very general or very specific. They may consist of a one-sided general note, or they may be laid down in a multiple page policy document. The second major difficulty is that it is not easy to ascertain that policies are really embedded within an organisation. In fact this should actually be made more concrete, for instance by adding characteristics of embedded policies such as: is published on the website; is updated on a regular basis, etc. A provisional approach, chosen here, is to add a line about the support of policies by the management of an organisation.

Example question (Core Survey 2, 14): *Does your organisation have a written digitisation strategy, endorsed by the management of your organisation? [Yes/No/Do not know]*

4. Degree of descriptive metadata cataloguing in a digital collection database

Explanatory note: This indicator essentially fixes the proportion of the collections of a heritage institution that are catalogued in a collection database.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Here the level at which cataloguing takes place is the main challenge to standardisation across the different heritage domains. In the *ENUMERATE* methodology it is attempted to overcome this by asking respondents to consider item level descriptions (metadata records) of analogue and born-digital heritage objects. Institutions where high level descriptions are common - in particular this is the case in archives and record offices - current cataloguing practices do not allow real item level cataloguing (yet).

Example question (Core Survey 2, 16): *Estimate the percentage of your entire heritage collection that has been catalogued in a collection database: ... [The estimated percentage of your entire heritage collections that has been catalogued in a collection database concerns item level descriptions (metadata records) of analogue and born-digital heritage objects.]*

5. Degree of digital reproduction/representation presence in the digital collections

Explanatory note: This indicator fixes the proportion of the analogue collections of a heritage institution that have been digitally reproduced.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: A digital reproduction is a digital surrogate of an original analogue object. In the *ENUMERATE* methodology an object that has only been catalogued in a database with metadata records is not considered to be “digitally reproduced”.

Example question (Core Survey 2, 17): *Estimate the percentage of your analogue heritage collections that has already been digitally reproduced: ... [A digital reproduction is a digital surrogate of an original analogue object. Please note that an object that has only been catalogued in a database with metadata records is not considered to be “digitally reproduced”.]*

6. Necessity to reproduce analogue heritage collections in digital format (per object type)

Explanatory note: This indicator focuses on the need to digitise that part of the analogue collections of an institution that have not been digitally reproduced yet, expressed as a percentage of the full analogue collections. The assumption is here that an institution may assess certain parts of the analogue collections as less necessary to be digitised. Of course this assessment may change over time.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: A digital reproduction is a digital surrogate of an original analogue object. In the *ENUMERATE* methodology an object that has only been catalogued in a database with metadata records is not considered to be “digitally reproduced”.

Example question (*Core Survey 2, 18*): *Estimate the percentage of your analogue heritage collections that still needs to be digitally reproduced: ... [A digital reproduction is a digital surrogate of an original analogue object. Please note that an object that has only been catalogued in a database with metadata records is not considered to be “digitally reproduced”.]*

Usage of digital material (demand)

7. Presence of institutional website(s)

Explanatory note: This indicator ascertains whether the heritage institution has a web presence, initiated and supervised by the institution.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: The concept of “institutional website” includes a website, home page, or presence on another entity’s website. It may but does not necessarily include an access point to the digital collections of an institution. Excluded are mere references to the institution in online directories and other web pages that are not under the control of the institution.

Example question (*Core Survey 2, 4*): *Please provide the address of your institution’s main website that is accessible for the general public.*

8. Importance of different types of use of digital collections

Explanatory note: This indicator measures the perceived importance of different types of use of digital heritage collections. It is specifically intended to trace changes in the focus on a limited classification of audiences. This is the classification proposed here:

- Academic research
- Creative reuse/Remix
- Educational use
- Ideological, religious and commemorative use
- Personal enjoyment

- Reducing the use of the physical originals
- Sales, commercial licensing
- Other types of use

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution.

Remarks: In actual surveys respondents should be invited to specify the class of *Other types of use*. An analysis of responses may result in an adaptation of the classification here above.

Example question (Core Survey 2, 19): *Collections are made accessible to the public for various reasons. How important is each of the following types of use for your institution? [Using a 10-points scale - where 1 equals "not at all important" to 10 "highly important" - please select only one number per row.]*

9. Existence of embedded policies related to the use of digital collections

Explanatory note: This indicator is a measure of the support for activities related to the use of an institution's digital collections.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: As for Indicator 3 above the *ENUMERATE* Team encountered that policies may be very general or very specific. Furthermore it is not easy to ascertain that such policies are really embedded within an organisation. Questions about this indicator may become more precise when "embedded policies" are defined as a formal policy document detailing which digital materials are accessible to whom and what the terms and conditions of this accessibility are. Refer to the notes to Indicator 3 for further details.

Example question (Core Survey 2, 20): *Does your organisation have an explicit (written) policy, endorsed by the management of your organisation, that sets conditions for specific types of use of your digital heritage collections, as specified for instance in the previous question? [Yes/No/Do not know]*

10. Status of monitoring the access to and use of digital collections

Explanatory note: This indicator establishes whether an institution monitors y/n the actual use (accessing) of its digital collections.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Any manner of measuring will suffice. However, the indicator should ideally exclude one-off monitoring. This can be done by adding a qualifier like: "on a regular basis". Once the proportion of institutions saying monitoring access is being put into practice approaches 100% (in a specific domain), it will not be a useful indicator anymore.

Example question (Core Survey 2, 21): *Does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users? [Yes/No/Do not know]*

11. Use of monitoring methods

Explanatory note: As an extension to Indicator 10, where the aim is to monitor the proportion of the institutions that put 'monitoring into practice', this is an indicator of the popularity of a limited number of monitoring methods:

- Website statistics
- Social media statistics (e.g. Facebook, Flickr, Youtube, Wikipedia)
- Database statistics (if not included in Website statistics and Social media statistics)
- User studies
- Other monitoring methods

An institution may of course make use of more than one of these methods.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: In actual surveys respondents should be invited to specify the class of *Other monitoring methods*. An analysis of responses may result in an adaptation of the classification here above.

Example question (*Core Survey 2, 22*): *How does your organisation measure the number of times digital metadata and/or digital objects are being accessed by your users? [Please indicate all ways in which the access of digital metadata and objects is measured.]*

12. Popularity of network access options regarding digital collections

Explanatory note: This indicator measures the estimated popularity of current and expected future access channels, starting from a predefined list of channels. It intends to trace over time changes in the popularity of these channels. To fix expectations about the popularity of access channels in the near future, the near future is defined as "two years later".

This is the classification proposed in this framework:

- Offline
- Institutional website
- National aggregator
- Europeana
- Other aggregator
- Wikipedia
- Other Social media platforms like Flickr, Youtube, Facebook
- Institutional API
- 3rd party API
- Other access channels

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution and size of institution.

Remarks: In actual surveys respondents should be invited to specify the class of *Other access channels*. An analysis of responses may result in an adaptation of the classification here above.

Example question (Core Survey 2, 23): *Please indicate estimated percentage of all the digital objects you have that are and/or will be accessible through the mentioned access options:... [per class of the access channels mentioned above two percentages were asked for: 1. Estimated percentage of digital objects currently accessible; 2. Estimated percentage of digital objects accessible 2 years from now. An estimation is OK in both cases.]*

Costs of digitisation (economics)

13. Annual expenditures on digital collections

Explanatory note: The indicator allows institutions to compare their institutional (internal) budgets and temporary (external) project budgets with these budgets in other institutions. On an aggregated scale it provides reference materials for the total cost of ownership of digital heritage collections.

Survey schedule: The preferred schedule is once a year (Feb-Apr). The indicator covers one full budget year. If the budget year for an institution does not coincide with the calendar, the amount should be related to the calendar year that fits best (in terms of the number of months). On pragmatic grounds a cycle of 2 to 3 years may be adopted, but the indicator is necessarily always about one specific year.

Disaggregation: Useful are disaggregation by type of institution and by size of institution.

Remarks: These budgets should be estimates of the costs related to the initial creation, on-going maintenance, enhancement and preservation of the digital collections of the responding institution. Included in these estimates should be the cost of the staff time devoted to digital collection related activities.

Example question (Core Survey 2, 26): *Please estimate your annual expenditure on your digital collections (total cost of ownership) [Institutional expenditure (internal budget) and temporary funded project expenditure (external budget) can be filled in separately. In both cases the respondent is asked: "Please specify the year concerned:"]*

14. Number of staff engaged in creating and preserving digital collections

Explanatory note: This indicator essentially fixes the number of paid staff in full-time equivalents that are engaged during the reference period (i.e. one year) in activities related to the initial creation, on-going maintenance, enhancement and preservation of the digital collections of an institution.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted, but the standardised indicator is always about one specific year.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Included should be the time of an institution's own staff engaged in activities related to creating/acquiring, maintaining, enhancing and preserving the digital collections, which includes: planning and managing in-house and contracted projects; preparing and digitising materials; and enhancing digitised output to widen accessibility.

Example question (Core Survey 2, 29): *What is the total number of paid staff (in full-time equivalent) engaged in creating/acquiring, maintaining, enhancing and preserving your *digital collections* on an annual basis?*

15. Sources of funding for digital collections

Explanatory note: High level, aggregated statistics are collected here. The indicator quantifies the prevalence of funding sources. This indicator has little immediate relevance to individual institutions. Measured is the percentage of institutions that make use of specific sources of funding. A limited classification is proposed for this indicator:

- Internal budgets
- Crowd funding
- National Public grant/subsidy
- Regional/Local Public grant/subsidy
- Private funds and legacies
- Public/private partnership
- Sales of digital items
- Other sources of funding

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution and size of institution.

Remarks: It is important to note that no actual figures are measured per funding source. An aggregated result does not show the relative volume of funding sources, it indicates the popularity of funding sources, per domain, country, etc. In an actual survey respondents should be invited to specify the class of *Other sources of funding*. An analysis of responses may result in an adaptation of the classification here above.

Example question (Core Survey 2, 31): *From what sources are your digital collection activities funded?: ... [respondents can check out all sources that are being used by their institution]*

16. Ratio of incidental versus structural costs of digital collections

Explanatory note: Costs of digital collections can be 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. Examples: selection of materials, acquisition of digital born materials, scanning, descriptive metadata creation, project management.
- *Structural costs* are the costs needed for the ongoing maintenance, enhancement and preservation of a digital collection. Examples: activities concerning the preservation of digital collections, licenses, maintenance of web servers, user outreach and support, management.

The indicator allows institutions to compare the ratio of incidental versus structural costs for their own institution with the ratio in other institutions. On an aggregated scale it will show the development of this ratio in a heritage domain, country, or the EU as a whole.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Useful are disaggregation by type of institution; by size of institution.

Remarks: It may be expected that over the years structural costs will become more significant.

Example question (Core Survey 2, 27): *Please estimate what percentage of the total annual expenditures on creating/acquiring, maintaining, enhancing and preserving your digital collections can be assigned to incidental costs and what percentage can be assigned to structural costs: ... [Percentages should add up to 100%]*

17. Ratio of in-house versus outsourced costs of digital collections

Explanatory note: The costs of digital collections can be divided into in-house costs and outsourced costs. The indicator allows institutions to compare the ratio of these costs for their own institution with the ratio in other institutions. On an aggregated scale it will show the development of this ratio in for instance a specific heritage domain, a country, or the EU as a whole.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Useful are disaggregation by type of institution; by size of institution.

Remarks: None.

Example question (Core Survey 2, 28): *Please estimate what percentage of the total annual expenditures on creating/acquiring, maintaining, enhancing and preserving your digital collections is spent In-house and what percentage is Outsourced: ... [Percentages should add up to 100%]*

18. Relative weight of costs associated with creating and preserving digital collections

Explanatory note: One of the most precarious topics in the work being done by *ENUMERATE* concerns the specification of costs. Since a cross-domain methodology to monitor the absolute costs of building up and maintaining digital collections is currently impossible - in particular because the operations related to digital heritage collections may vary immensely -, the approach chosen here is asking respondents to assess the relative weight of the main activities in building up and maintaining digital collections. This is done for both the Incidental costs and the Structural costs (refer to Indicator 16). The following activity-based costs were defined:

Incidental cost categories:

- Project management
- Selection of material for digitisation
- Acquisition of digital born material
- Logistics (shipment of collection for digitisation, etc.)
- Analogue to 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
- Web design, software development
- Other (Incidental) costs

Structural cost categories:

- Management
- Archiving (storage, including backups)
- Activities concerning the long-term preservation of the digital collection (including research activities but excluding Archiving costs)
- Licences
- Maintenance of web servers and web, mobile and other services
- User outreach and support (including staff time for efforts to promote the use of the digital collections)
- Usage analysis (including user surveys, interviews, and other activities)
- Editorial (including content selection and updating)
- Other (Structural) costs

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Useful are disaggregation by type of institution.

Remarks: In Core Survey 2 respondents were free to choose whether they would answer this question. In an actual survey respondents may be invited to specify the class of *Other (Incidental/Structural) costs*. An analysis of responses may result in an adaptation of the classification.

Example question (Core Survey 2, 28+): *I. Please estimate what percentages of the Incidental costs can be assigned to the following activities: ... II. Please estimate what percentages of the Structural costs can be assigned to the following activities: ... [In both cases percentages should add up to 100%. The costs are specified in the Explanatory note above.]*

Digital preservation practices (sustainability)

19. Existence of embedded of policies related to the sustainability of digital collections

Explanatory note: This indicator is a measure of the internal/managerial support for activities related to the preservation of an institution's digital collections.

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: As for Indicators 3 and 9 above the *ENUMERATE* Team acknowledges that policies may be very general or very specific. It is not easy to ascertain that policies are really embedded within an organisation. Questions about this indicator will become more concrete when "embedded policies" are defined as a formal policy document describing the strategy for the digital preservation and permanent access to the digital collections.

Example question (Core Survey 2, 24): *Does your organisation have a written Digital Preservation Strategy that is endorsed by the management of your organisation? [Yes/No/Do not know]*

20. Adherence to international standards in digital preservation practice

Explanatory note: Adherence to international standards in digital preservation practice is understood as a measure of achieved maturity in making the digital collections of an institution sustainable.

For the *ENUMERATE* framework three classes of digital archives have been defined:

- Institutional digital archives that meet the international criteria for long term preservation
- Publicly managed professional digital archives
- Privately managed professional digital archives

(The assumption is here that a 'professional digital archive' is set up according to international criteria for long term preservation.)

Survey schedule: The preferred schedule is once a year (Feb-Apr). On pragmatic grounds a cycle of 2 to 3 years may be adopted.

Disaggregation: Suggested is disaggregation by type of institution; by size of institution.

Remarks: Since the *ENUMERATE* Team found time and again that only a minor fraction of the memory institutions covered by this framework is (as of 2013) successful in setting up or making use of professional digital archives, the percentage of institutions that will answer in the affirmative is expected to be low for some time to come. In Core Survey 2 a definition of ‘professional digital archive’ was not given.

Example question (Core Survey 2, 25): *Are your digital collections stored in digital archives that have been set up according to *international standards* for *digital preservation*? [The following 5 answering options were presented:*

- *Yes, we have our own digital archive that meets the international criteria for long term preservation*
- *Yes, our digital collections are archived in a publicly managed professional digital archive*
- *Yes, our digital collections are archived in a privately managed professional digital archive*
- *No, we do not have a solution yet for the long term preservation of our digital collections based on international standards*
- *Do not know*

5.2 Classification of heritage materials

The digital collection consists of digitally reproduced analogue objects and born digital objects. An object that has been catalogued in a database with metadata records **only**, is not considered to be part of the ‘digital collection’.

Collection type	Object type
(01) TEXT BASED RESOURCES	
	Rare printed books
	Other printed books
	Electronic books (eBooks)
	Newspapers
	Journals
	Other serials
	Medieval Manuscripts
	Other Manuscripts
	Microforms and microfilms
	Other text based
(02) VISUAL (2D) RESOURCES	
	Drawings
	Engravings / Prints
	Maps and ground plans
	Paintings

	Photographs
	Posters
	Sheet music
	Other visual resources
(03) ARCHIVAL RESOURCES (not included in 01 or 02)	
	Archives : Government documents
	Archives : Other archival records
(04) 3D MAN-MADE MOVABLE OBJECTS	
	3 Dimensional works of art
	Archaeological Furnishings and Equipment
	Other Furnishings and Equipment
	Coins and medals
	Other 3 dimensional man-made objects
(05) NATURAL RESOURCES	
	Natural inert specimens
	Natural living specimens
(06) GEOGRAPHY BASED RESOURCES	
	Monuments and buildings
	Landscapes
	Archaeological sites
	Other geography based resources
(07) TIME BASED RESOURCES	
	Audio files: Music
	Audio files: Speech & other (excluding digital audio books; including oral history)
	Digital audio books
	Film
	Video recordings
	Other time based resources
(08) DIGITAL INTERACTIVE RESOURCES (EXCLUSIVELY DIGITAL)	
	Databases (containing cultural heritage metadata)
	Digital (3D) designs or reconstructions of objects and buildings

D2.12 – Conceptual Framework

	Born-digital art objects
	Digital research files (including GIS files)
	Games
	Software (customised)
	Websites (and parts of websites)
	Other born-digital interactive resources

5.3 The *ENUMERATE* Definitions

Term	Definition
3D movable objects	Tangible objects - having or appearing to have, the three dimensions of length, width, and height - capable of being moved or conveyed from one place to another, as opposed to real estate or other buildings
Access	Right, opportunity or means of obtaining information from documents.
Access control	Verification of user rights and the terms and conditions for the access to a publication.
Access policy	A formal written statement issued by the person(s) or body responsible for managing archives or special collections, specifying which materials are available for access and by whom, including any conditions or restrictions on use, usually posted or distributed by some method to users.
API	Stands for "Application Program Interface," though it is sometimes referred to as an "Application Programming Interface." An API is a set of commands, functions, and protocols which programmers can use when building software for a specific operating system. The API allows programmers to use predefined functions to interact with the operating system, instead of writing them from scratch.
Archaeological man-made objects	Objects that are made by humans rather than being natural, found in contexts of archaeological research, more specifically, those unearthed from periods before usually the 19th century and from extinct cultures.
Archival records	Documents created or received and maintained by an agency, organization or individual in pursuance of legal obligations, in the transaction of business, or in the course of the conduct of affairs, and preserved because of their enduring value.
Archival resources	Resources, regardless of form or medium, created and/or accumulated and used by a particular person, family, or corporate body in the course of that creator's activities and functions.
Archive	The division within an organization responsible for maintaining the organisation's records of enduring value. An organisation that collects the records of individuals, families, or other organisations.
Article	Independent text forming a part of a publication
Audio book/Digital audio book	A (digital) audio book is a (digital) recording of a text being read, e.g. on CD or DVD, or by means of other (digital) technologies.
Audio recording	Any medium on which sounds are recorded for mechanical or electronic playback, including phonograph records (vinyl), audiotape, and compact disc. Synonymous with sound recording.
Audio visual document	Document in which sound and/or pictures are prominent, and which requires the use of special equipment to be seen and/or heard.
Book	Non-serial printed document in codex form.
Born digital/ Born digital heritage	Digital materials which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form.

Buildings	Structures, generally enclosed, that are used or intended to be used for sheltering an activity or occupancy.
Coins	Pieces of metal stamped by government authority for use as money.
Collection	Gathering of documents assembled on the basis of some common characteristic, without regard to their provenance.
Collection registration system (CRS)	Automated system with the purpose of recording information about objects in a collection.
Creative reuse	Reuse with the aim of making or composing a (new) creative product.
Cultural heritage institutions	Memory institutions (museums, libraries, archives/records offices, audio-visual and film institutes, institutes with curatorial care for monuments and sites, and some other hybrid types of organisations). The criterion is that curatorial care is for part of the collections of the institution that are included in its mission. Institutions that have collections of heritage materials (like for example of books, films, and music) to be lent or sold to contemporary users without the explicit task of safeguarding the collections for future generations, are excluded. This essentially leaves out both school libraries and public libraries without cultural heritage collections.
Databases	Structured sets of data held in computer storage, especially those that incorporate software to make them accessible in a variety of ways. A database is used to store, query, and retrieve information, typically comprising a logical collection of interrelated information that is managed as a unit, stored in machine-readable form, and organized and structured as records that are presented in a standardized format in order to allow rapid search and retrieval by a computer.
Digital art	Digital art is a general term for a range of artistic works and practices that use digital technology as an essential part of the creative and/or presentation process.
Digital collections	The whole of digitised and born-digital cultural objects and their metadata within a cultural institution.
Digital interactive resources	Digital interactive resources are products and services on computer-based (digital) systems which respond to the user's actions by presenting (response eliciting) content such as text, graphics, animation, video, audio, and games.
Digital materials	A broad term encompassing digital surrogates created as a result of converting analogue materials to digital form (digitisation), and "born digital" for which there has never been and is never intended to be an analogue equivalent, and digital records.
Digital preservation	The process of maintaining, in a condition suitable for use, materials produced in digital formats, including preservation of the bit stream and the continued ability to render or display the content represented by the bit stream. The task is compounded by the fact that some digital storage media deteriorate quickly ("bit rot"), and the digital object is inextricably entwined with its access environment (software and hardware), which is evolving in a continuous cycle of innovation and obsolescence. Also refers to the practice of digitizing materials originally produced in non-digital formats (e.g. print and film) to prevent permanent loss due to deterioration of the physical medium.
Digital preservation infrastructure	The basic physical and organisational structures and facilities (e.g., hardware, software, system management facilities, etc.) needed for the implementation of digital preservation.

Digital preservation strategy (written)	A document formally approved within an organisation describing the way the organisation will be active in the preservation of her digitized and born digital collections.
Digital research files	Structured collections of scientific or scholarly data, or resources for storing such data, which can be processed by a computer program and are which usually based on some kind of durable storage.
Digitisation	The process of converting, creating and maintaining books, art works, historical documents, photos, journals etc, in electronic representation so they can be viewed via computer and other devices.
Digitisation strategy (written)	A document formally approved within an organisation describing the way the organisation will be active in the digitisation of her collections.
Document	Recorded information or material object, which can be treated as a unit in a documentation process.
Download	In computer networks, to download means to receive data to a local system from a remote system or to initiate such a data transfer. Examples of a remote system from which a download might be performed include a web server, FTP server, email server, or other similar systems. A download can mean either any file that is offered for downloading or that has been downloaded, or the process of receiving such a file. It has become more common to mistake and confuse the meaning of downloading and installing or simply combine them incorrectly together.
Drawing	Picture made with a solid mineral substance or a pointed tool.
Electronic books (eBooks)	A book composed in or converted to digital format for display on a computer screen or handheld device.
Engraving	Print made from any kind of intaglio plate, whether engraved with hand-tools or a machine, or etched with acid, so that the printing areas are lower than the non-printing areas.
Film	Series of pictures recorded on a strip of transparent material, or on an electronic data medium, which, when projected or produced rapidly one after another on a screen, give the illusion of natural and continuous movement.
Free commercial downloads, sharing and reuse	One is free to download, share (copy and redistribute) the material in any medium or format and adapt (remix, transform, and build upon) the material for any purpose, even commercially.
Free non-commercial downloads, sharing and reuse	One is free to share (copy and redistribute) the material in any medium or format and adapt (remix, transform, and build upon) the material, but one may not use the material for commercial purposes.
Free restricted access	Free access to the material but with certain limitations (for example, a registration is required to gain access).
Full-time equivalent (FTE)	A measurement equal to one staff person working a full-time work schedule for one year. Example If out of three persons employed in an institution, one works quarter-time, one works half-time, and one works full-time, then the FTE of these three persons would be $0,25 + 0,5 + 1,0 = 1,75$ employees (FTE).
Games	Forms of competitive play, usually involving an element of strategy, especially to interfere with an opponent's play; may or may not require physical skill.

Geo-blocking	Restricting access only for users from a certain geographical region (e.g. a country).
Geography based resources	Objects (monuments, buildings, landscapes, etc.) that are inherently linked to a specific location or geographical area on the surface of the Earth and can be referenced in the form of geographic coordinates and/or geographic names.
Government archives	The whole of the records made or received by an agency of government and maintained in the conduct of government business.
Government documents	Records made or received by an agency of government and maintained in the conduct of government business.
Ground plan	Drawings, sketches or prints, depicting a building or any object viewed from above, geometrically represented as projected on a horizontal plane.
Incidental cost (of digital collections)	Costs associated with the creation or acquisition of (new) digital collections. These may include the costs of digitisation, metadata creation, project management, IP rights clearance, user experience research and usage analysis, website design, programming, preservation, and outreach efforts.
Institutional website	A website that is ran by or commissioned by the institution. This can be the official homepage of the institution, but may also be a thematic website.
Issues	Copy of a newspaper or periodical published on the same date and bearing the same issue number
Item-level metadata	The bibliographic and descriptive metadata needed to include item level records in your online catalogue.
Journal	Serial under the same title published at regular or irregular intervals, over an indefinite period, individual issues in the series being numbered consecutively or each issue being dated. NOTE: Series of reports, transactions of institutions, series of regular conference proceedings and annuals are included, while newspapers and monographic series are excluded.
Key Performance Indicators (KPI)	Key Performance Indicators (KPIs) are trends or facts that are commonly used by an organisation to evaluate its success or the success of a particular activity in which it is engaged.
Landscapes	Broadly used to describe portions of the earth's surface that share common repeating characteristics that can be comprehended at a glance. Landscapes are more than scenery or political units; they are systems of natural and cultural contexts.
Library	Organisation or part of an organisation, the main aims of which are to build and maintain a collection and to facilitate the use of such information resources and facilities as are required to meet the informational, research, educational, cultural or recreational needs of its users. Note: These are the basic requirements for a library and do not exclude any additional resources and services incidental to its main purpose.
Manuscript	Original document that is handwritten or in typescript. NOTE: Bound volumes and other units (fragments, rolls, autographs, etc.) may be counted separately.
Map	Conventional representation, on a reduced scale and usually flat, of phenomena which can be localized in space and time.

Medals	Small pieces of metal, usually gold, silver, or bronze, bearing a relief design on one or both sides and having a commemorative purpose; not used as a medium of exchange.
Medieval Manuscript	Manuscript original document that is handwritten or in typescript in medieval times.
Metadata	“Data about data.” Structured information describing information resources/objects for a variety of purposes....The term is generally used in the library community for non-traditional schemes such as the Dublin Core Metadata Element Set, the VRA Core Categories, and the Encoded Archival Description (EAD). Metadata has been categorized as descriptive, structural, and administrative. Descriptive metadata facilitates indexing, discovery, identification, and selection. Structural metadata describes the internal structure of complex information resources. Administrative metadata aids in the management of resources and may include rights management metadata, preservation metadata, and technical metadata describing the physical characteristics of a resource.
Microform	Photographic document requiring magnification when used. NOTE 1: Microfiche and microfilm are included. NOTE 2: Slides and similar documents are counted as audiovisual documents.
Monument	Historic monuments are fixed assets that are identifiable because of particular historic, national, regional, local, religious or symbolic significance. This includes architectural works, groups of buildings, works of monumental sculpture and painting, elements or structures of an archaeological nature, inscriptions, cave dwellings and combinations of features. The definition excludes objects in the collections of archives, libraries and museums.
Museum (1)	Organized collection of artefacts or naturalia of cultural or scientific interest, stored permanently for intended display.
Museum (2)	Organization or part of an organization responsible for collecting, preserving, and exhibiting museum documents.
Music	Combined vocal or instrumental sounds in measured time to communicate emotions, ideas, or states of mind, usually according to cultural standards of rhythm, melody, and, in most Western music, harmony.
National aggregator	A website that aggregates the digital collections of a nation's heritage institutions and serves as an access point to these collections.
National digital preservation strategy	An official, national or federal, cross-institutional policy on the preservation of digitized cultural heritage and born digital cultural heritage that is imposed upon the institutions that are included in the strategy. Institutions may take part in the formative process of such a strategy and, as such, have a role in enforcing the strategy.
National digitisation strategy	An official, national or federal, cross-institutional policy on the digitization of cultural heritage that is imposed upon the institutions that is included in the strategy. Institutions may take part in the formative process of such a strategy and, as such, have a role in enforcing the strategy.
Natural resources	Commodities or assets with some economic value that exist without any effort of mankind. Natural resources can be of three types. The first is non-renewable, like oil and coal, stocks of which will eventually run out. The second is renewable, like water and fish, which are reproducible. The third is non-expendable: it is not used up in the consumption process. An example is a landscape of outstanding beauty, which yields utility for those

	that see it and tourist income for the owner.
Newspaper	Serial, which contains news on current events of special or general interest, the individual parts of which are listed chronologically or numerically and usually appear at least once a week.
Non-restricted access	Access to the material without any limitations.
Offline	A digital collection that is available on the local area network of the institution and which cannot be accessed through the internet.
Ongoing costs	Costs of maintaining, enhancing, and preserving (...) collections that have already been digitized. These costs may include adding new content to an existing collection or updating previously digitized content, new metadata creation, project management, IP rights clearance, user experience research and usage analysis, website design, programming, preservation, and outreach efforts. This may include all those costs incurred once the digitized collection has been launched.
Paid access for downloads, sharing and reuse	After paying a fee, one is free to download, share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material).
Paid restricted access	Access to the material after paying a fee, and with additional limitations (for example, a registration is required to gain access).
Painting	A work produced through the art of painting in oil, acrylic paint, watercolour
Photograph	Picture obtained by a process which fixes a direct and durable image on a sensitized surface by the action of electromagnetic radiation.
Platform	A computing platform includes hardware architecture and a software framework, including application frameworks, where the combination allows software, particularly application software, to run. Typical platforms include a computer architecture, operating system, programming languages and related user interface (run-time system libraries or graphical user interface).
Postcard	Card for conveyance by post, often with a picture on one side.
Poster	A large single sheet of heavy paper or cardboard, usually printed on one side only, with or without illustration, to advertise a product/service or publicize a forthcoming event (meeting, concert, dramatic performance, etc.), intended for display on a bulletin board, kiosk, wall, or other suitable surface.
Preservation	All measures taken including financial and strategic decisions, to maintain the integrity and extend the life of documents or collections.
Print	Copy of an image transferred to a sensitive material.
Rare book	Book published before 1800.
Rare printed book	Book published before 1800.
Record	Document created or received and maintained by an agency, organization or individual in pursuance of legal obligations or in the transaction of

	business.
Remix	In the context of the <i>ENUMERATE</i> surveys a remix is a digital object that has been edited in order to be combined with other digital objects or to be experienced as an artistic product that is different from the original version.
Reuse	Reuse is defined (here) as: ...the use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced.
Serial	A publication in any medium issued under the same title in a succession of discrete parts, usually numbered, or dated, and appearing at regular or irregular intervals with no predetermined conclusion.
Sheet music	Printed music issued without covers, whether actually printed on single sheets (pages) or not.
Software	Organised collections of computer data and instructions.
Speech	Expressed or described thoughts, feelings, or perceptions by the articulation of words.
Structural cost (of digital collections)	The sum total of the annual costs needed for the ongoing maintenance, enhancement and preservation of a digital collection. These costs may include adding new content to an existing collection or updating previously digitised content, new metadata creation, project management, IP rights clearance, user experience research and usage analysis, website design, programming, preservation costs, and outreach efforts.
Text based resources	Resources consisting of written or printed words, phrases, or sentences arranged to make a communication.
Thematic aggregator	A website that aggregates digital collections that are thematically linked and which serves as an access point to these collections.
Three-dimensional works of art	A work of art that has height, width and depth.
Time based resources	Time based resources are resources of (cultural heritage) objects that are not static but evolve during a certain time frame like video, audio, film, etc.
Upfront costs	Costs associated with the creation of new digitized (...) collections. These may include the costs of digitization, metadata creation, project management, IP rights clearance, user experience research and usage analysis, website design, programming, preservation, and outreach efforts. This may include all those costs needed to initially create the resource.
Usage	The act of frequenting a particular website by its visitors.
User statistics	The statistics that capture the degree in which users are frequenting a specific website.
Video recording	Electronic medium in which visual images, usually in motion and accompanied by sound, are recorded for playback by means of a television receiver or monitor.
Visitors	The number of different persons (users) that accessed a particular website during the reporting period. A visitor is defined as a 'unique visitor' during the reporting period only. A usual reporting period is one month. Someone visiting a website in five separate months is counted five times as a unique visitor, although it concerns one and the same person. Therefore it is methodologically incorrect to compute the yearly sum total of unique

	visitors based on monthly data. By definition the number of unique visitors can never be higher than the number of visits.
Visits	A series of consecutive user actions within a fixed period of time. Usually a visit is defined as being completed after thirty minutes of non-activity. Different definitions (e.g. other lengths of the time period) will result in different statistics. By definition the number of visits can never be higher than the number of pages visited.
Visual resources	Any items that communicate information primarily visually and are collected and made available as information resources.
Web analytics	Is the measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing web usage.
Websites	Sets of related web pages served from a single web domain. Websites are hosted on web servers, and are accessible via a network such as the Internet or private local area networks.

5.4 The *ENUMERATE* Core Survey 2 questionnaire



ENUMERATE Core Survey 2 October 2013

Dear colleague,

On behalf of the community of archives, libraries and museums we ask you to help us by participating in this survey. There is a growing demand for reliable data about:

- Digitisation activity
- The cost of digital collections
- Access to digital collections
- The preservation of digital heritage materials

This survey is a follow-up of the successful *ENUMERATE* Core Survey of 2012, and is designed to inform both your own institutional policies and policies on a national and European level. Together we can make the case for investments in our digital activities.

The survey is organised by the EU-funded *ENUMERATE* network. This is a community of practice in the field of digital cultural heritage. The survey is distributed among thousands of institutions across Europe.

The data collected will be kept strictly anonymous. The information that you share with us will not be published in a way that is traceable to your institution. The data is primarily used to establish a statistical measure of the progress of digitisation in European countries.

The survey consists of **33 questions**. The time needed to answer these will depend on the availability of management information about your digitisation activities. If the information is not readily available additional research in your institution may be needed. The reward for such efforts will go both to your institution and the community of memory institutions in Europe.

We hope you will be able to complete the questionnaire **by November 15th 2013**. You can pause and return at any time to continue. The meaning of words marked with asterisks can be looked up in the list of definitions (upper right hand corner).

Please do not hesitate to contact us through the e-mail address listed below if you need more information about the questions or the survey in general.

We thank you in advance for your kind attention and participation.

Best Regards, the *ENUMERATE* Team

ENUMERATE@panteia.nl

You can get a downloadable copy of the questionnaire <<here>>

More information on the *ENUMERATE* project and its network is available on www.enumerate.eu. The report containing key findings of the 2012 Core Survey can be downloaded from www.enumerate.eu/en/statistics/

SECTION 1/7: Organisational Information

1. Name of institution / organisation

[Input box]

This information will not be published in the *ENUMERATE* report.

2. Type / Domain of institution / organisation

Specify the primary heading you would assign to your institution.

Please choose **only one** of the following:

- National archive
- Other archive / records office
- Audio-visual / broadcasting archive
- Film institute
- Institution for performing arts
- Museum of art
- Museum of archaeology or history
- Museum of natural history or natural science
- Museum of science or technology
- Museum of ethnography or anthropology
- National library
- Higher education library
- Public library
- Special or other type of library
- Institution for monument care
- Other (specify below)

Other type of institution: [input box]

3. Country in which your institution is located

- Albania
- Andorra
- Armenia
- Austria
- Azerbaijan
- Belgium
- Bosnia and Herzegovina
- Bulgaria
- Croatia
- Cyprus
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Latvia
- Liechtenstein

- Lithuania
- Luxembourg
- Malta
- Republic of Moldova
- Monaco
- Montenegro
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Russian Federation
- San Marino
- Serbia
- Slovak Republic
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- Ukraine
- United Kingdom
- Former Yugoslav Republic of Macedonia
- Other country (specify below)

Other country: [input box]

4. Website of your institution

[Input box]

Provide the address of your institution's main website that is accessible for the general public.

This information will not be published in the *ENUMERATE* report.

5. Your name

[Input box]

This information will not be published in the *ENUMERATE* report.

6. Your job title

[Input box]

The role or position of the main person completing this survey.

7. Your e-mail address

[Input box]

This information will not be published in the *ENUMERATE* report.

8. Your telephone number and/or Skype contact details

[Input box]

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

This information will not be published in the *ENUMERATE* report.

9. What is your institution's total annual budget?

Please choose **only one** of the following:

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

Provide the annual budget for the entire cultural heritage institution as indicated in the last published annual account. If your institution is part of a larger organisation (e.g. a higher education library that is part of a higher education institution) only provide the budget of the cultural heritage related unit.

The total annual budget may include government funding, project funding, revenues from commercial activities, etc. If your budget occurs in two categories (e.g. €50,000), please choose the lower category.

10. Total number of paid staff (in *full time equivalents*, not in number of people)

[Input box] [Only 1 decimal accepted, e.g.: 3,7]

The number of *full time equivalents* should represent the total staff employed by your institution, including permanent and temporary staff, but excluding contractors and volunteers. Part-time staff needs to be added up to represent a full working week.

Note: the number of staff engaged in *digitisation activities* will be asked for later in the survey (see below).

SECTION 2/7: Digitisation Activity

11. Does your institution have *collections* that need to be preserved for future generations?

Yes

No

Answer this question with 'No' if your institution does not hold heritage collections or if you only have collections (for example of books, films, music) that can be lent or sold to users.

If the answer is No, automatically proceed to o-o-o at the end of the survey.

12. Does your organisation have *digital collections* or is it currently involved in collection *digitisation* activities?

Yes

No

If the answer is No, automatically proceed to questions 13, 14, 15, 16, 17, 18 and after that to the end of the survey.

13. Does your organisation have a *written digitisation strategy*, endorsed by the management of your organisation?

Yes

No

Do not know

The strategy may be for any period up to 2020.

14. Does your organisation collect *born digital heritage*?

Yes

No

Do not know

Answer this question with 'yes' if your institution collects any kind of *born digital heritage* materials (i.e. software, digital documents, digital art, harvested web content, etc.) with the explicit intention of preserving these born digital materials for future generations.

15. Please select the collection types that are part of the heritage collections of your institution

NOTE: Table will not be presented as it is here. The table can be folded out selectively, starting from the high level collection type classes in the left column.

Please specify the object types that are part of the heritage collections of your institution. The digital collection consists of digitally reproduced analogue objects and born digital objects. An object that has been catalogued in a database with metadata records **only**, is not considered to be part of the 'digital collection'.

Collection type	Object type	In analogue collection y/n	In digital collection y/n
(01) TEXT BASED RESOURCES			
	Rare printed books		
	Other printed books		
	Electronic books (eBooks)		
	Newspapers		
	Journals		
	Other serials		
	Medieval Manuscripts		
	Other Manuscripts		
	Microforms and microfilms		
	Other text based		
(02) VISUAL (2D) RESOURCES			
	Drawings		
	Engravings / Prints		

	Maps and ground plans		
	Paintings		
	Photographs		
	Posters		
	Sheet music		
	Other visual resources		
(03) ARCHIVAL RESOURCES (not included in 01 or 02)			
	Archives : Government documents		
	Archives : Other archival records		
	Archives : Other archives		
(04) 3D MAN-MADE MOVABLE OBJECTS			
	3 Dimensional works of art		
	Archaeological Furnishings and Equipment		
	Other Furnishings and Equipment		
	Coins and medals		
	Other 3 dimensional man-made objects		
(05) NATURAL RESOURCES			
	Natural inert specimens		
	Natural living specimens		
(06) GEOGRAPHY BASED RESOURCES			
	Monuments and buildings		
	Landscapes		
	Archaeological sites		
	Other geography based resources		
(07) TIME BASED RESOURCES			
	Audio files: Music		
	Audio files: Speech & other (excluding digital audio books; including oral history files)		
	Digital audio books		
	Film		
	Video recordings		
	Other time based resources		

(08) DIGITAL INTERACTIVE RESOURCES (EXCLUSIVELY DIGITAL)			
	Databases (containing cultural heritage metadata)		
	Digital (3D) designs or reconstructions of objects and buildings		
	Born-digital art objects		
	Digital research files (incl. GIS files)		
	Games		
	Software (customised)		
	Websites (and parts of websites)		
	Other born-digital interactive resources		

Tests have indicated that answering the questions about the size of collections are rewarding, but may be challenging. Some institutions will want to be more specific than is possible here; others may find it difficult to give even the high level estimates asked for. We are convinced that all institutions will benefit from an exercise in mapping out digital collections. Please consider the OPTIONAL Question below question #18 if you want to take up the challenge!

16. Estimate the percentage of your entire heritage collection that has been catalogued in a collection database

[Input box]

The estimated percentage of your entire heritage collections that has been catalogued in a collection database concerns item level descriptions (metadata records) of analogue and born-digital heritage objects.

17. Estimate the percentage of your analogue heritage collections that has already been digitally reproduced

[Input box]

A digital reproduction is a digital surrogate of an original analogue object. Please note that an object that has only been catalogued in a database with metadata records is not considered to be "digitally reproduced".

18. Estimate the percentage of your analogue heritage collections that still needs to be digitally reproduced

[Input box]

If it is difficult to provide global estimates of the size of your collections, please consider to map out the size of your collections per *object type*. The information gathered in the table below will be highly valuable both for your own institution and the international community of archives, libraries and museums.

[] Show table

Please indicate the size of your collections per object type and assess the need to digitise

NOTE: Table will not be presented as it is here. The table can be folded out selectively, starting from the high level collection type classes in the left column.

Collection type	(A) Number of records in your collection database(s) on item level	(B) Estimated number of analogue items in * <u>heritage collection</u> * [indicate units; see Core Survey 1]	(C) Estimated number of born digital items in * <u>heritage collection</u> *	(D) Estimated % of objects in <u>analogue collection</u> that is digitally reproduced	(E) Estimated % of objects in <u>analogue collection</u> still to be digitally reproduced	(F) Estimated % of objects in <u>analogue collection</u> with no need to be digitally reproduced
(01) TEXT BASED RESOURCES						
Rare printed books	... records	... volumes	not applicable	... %	... %	... %
Other printed books	... records	... volumes	not applicable	... %	... %	... %
Electronic books (eBooks)	... records	... volumes	... volumes	... %	... %	... %
Newspapers	... records	... issues	... issues	... %	... %	... %
Journals	... records	... issues	... issues	... %	... %	... %
Other serials	... records	... issues	... issues	... %	... %	... %
Medieval Manuscripts	... records	... items	not applicable	... %	... %	... %
Other Manuscripts	... records	... items	... items	... %	... %	... %
Microforms and microfilms	... records	... items	not applicable	... %	... %	... %
Other text based	... records	... items	... items	... %	... %	... %
(02) VISUAL (2D) RESOURCES						
Drawings	... records	... items	... items	... %	... %	... %
Engravings / Prints	... records	... items	not applicable	... %	... %	... %
Maps and ground plans	... records	... items	... items	... %	... %	... %
Paintings	... records	... items	... items	... %	... %	... %
Photographs	... records	... items	... items	... %	... %	... %
Posters %	... %	... %

	records	items	items			
Sheet music	... records	... items	... items	... %	... %	... %
Other visual resources	... records	... items	... items	... %	... %	... %
(03) ARCHIVAL RESOURCES (not included in 01 or 02)						
Archives : Government documents	... records	... metres	... records	... %	... %	... %
Archives : Other archival records	... records	... metres	... records	... %	... %	... %
Archives : Other archives	... records	... archives	... archives	... %	... %	... %
(04) 3D MAN-MADE MOVABLE OBJECTS						
3 Dimensional works of art	... records	... items	not applicable	... %	... %	... %
Archaeological Furnishings and Equipment	... records	... items	not applicable	... %	... %	... %
Other Furnishings and Equipment	... records	... items	not applicable	... %	... %	... %
Coins and medals	... records	... items	not applicable	... %	... %	... %
Other 3 dimensional man-made objects	... records	... items	not applicable	... %	... %	... %
(05) NATURAL RESOURCES						
Natural inert specimens	... records	... items	not applicable	... %	... %	... %
Natural living specimens	... records	... items	not applicable	... %	... %	... %
(06) GEOGRAPHY BASED RESOURCES						
Landscapes	... records	... landscapes	not applicable	... %	... %	... %
Archaeological sites	... records	... sites	not applicable	... %	... %	... %
Monuments and buildings	... records	... items	not applicable	... %	... %	... %

Other geography based resources	... records	... items	not applicable	... %	... %	... %
(07) TIME BASED RESOURCES						
Audio files: Music	... records	... hours	... hours	... %	... %	... %
Audio files: Speech & other (excluding digital audio books; including oral history files)	... records	... hours	... hours	... %	... %	... %
Digital audio books	... records	... volumes	... volumes	... %	... %	... %
Film	... records	... hours	... hours	... %	... %	... %
Video recordings	... records	... hours	... hours	... %	... %	... %
Other time based resources	... records	... items	... items	... %	... %	... %
(08) DIGITAL INTERACTIVE RESOURCES (EXCLUSIVELY DIGITAL)						
Databases (containing cultural heritage metadata)	... records	not applicable	... items	not applicable	not applicable	not applicable
Digital (3D) designs or reconstructions of objects and buildings	... records	not applicable	... items	not applicable	not applicable	not applicable
Born-digital art objects	... records	not applicable	... items	not applicable	not applicable	not applicable
Digital research files (incl. GIS files)	... records	not applicable	... items	not applicable	not applicable	not applicable
Games	... records	not applicable	... items	not applicable	not applicable	not applicable
Software (customised)	... records	not applicable	... items	not applicable	not applicable	not applicable
Websites (and parts of websites)	... records	not applicable	... items	not applicable	not applicable	not applicable
Other born-	...	not	...	not	not	not

digital interactive resources	records	applicable	items	applicable	applicable	applicable
-------------------------------	---------	------------	-------	------------	------------	------------

Explanatory notes for each of the 6 columns:

(A) Number of records in your collection database(s) on item level:

The (estimated) quantities of item level records in your collection database(s). (A 'record' is the container of the bibliographic and/or descriptive metadata per collection item.)

(B) Estimated number of analogue items in *heritage collection*

Provide the estimated number of analogue items that are part of your institutions heritage collections.

(C) Estimated number of born digital items in *heritage collection*:

Provide the estimated number of born digital items that are part of your institutions heritage collections.

(D) Estimated % of objects in analogue collection that is digitally reproduced:

You may include textual materials that are reproduced as images without character recognition.

Please include the digital output from projects that are approaching completion.

(E) Estimated % of objects in analogue collection still to be digitally reproduced:

The part of your analogue collection that your institution intends to digitise.

(F) Estimated % of objects in analogue collection with no need to be digitally reproduced:

The part of your collection that, for whatever reason, your institution does not intend to digitise, for instance because there is no demand for it or because it will be digitised by another institution.

SECTION 3/7. Digital Access

19. Collections are made accessible to the public for various reasons. How important is each of the following types of use for your institution?

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

Type of use	1	2	3	4	5	6	7	8	9	10
Academic research										
Creative reuse/Remix										
Educational use										
Ideological, religious and commemorative use										
Personal enjoyment										
Reducing the use of the physical originals										
Sales, commercial licensing										
Other types of use (specify below)										

Other types of use: [input box]

20. Does your organisation have an explicit (written) policy, endorsed by the management of your organisation, that sets conditions for specific types of use of your digital heritage collections, as specified for instance in the previous question?

Yes

No

Do not know

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

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

Yes

No

Do not know

In order to be able to answer this question with 'yes' any manner of measurement will suffice.

22. If Yes, how?

Website statistics

Social media statistics (e.g. Facebook, Flickr, Youtube, Wikipedia)

Database statistics (if not included in Website statistics and Social media statistics)

User studies

Other: [input box]

Please indicate all ways in which the access of digital metadata and objects is measured.

23. Please indicate estimated percentage of all the digital objects you have that are and/or will be accessible through the mentioned access options

Access channel	% of digital objects currently accessible (estimation is OK)	% of digital objects accessible 2 years from now (estimation is OK)
Offline		
Institutional website		
National aggregator		
Europeana		
Other aggregator		
Wikipedia		
Other *Social media platforms* like Flickr, Youtube, Facebook		
Institutional *API*		

3 rd party *API*		
Other Access channels (specify below)		

Other Access channels: [input box]

Multiple access options for your individual digital collections are a possibility (i.e. Europeana and Wikipedia). Consequently, the sum total of your answers does not have to be 100%.

SECTION 4/7. Digital Preservation

24. Does your organisation have a **written Digital Preservation Strategy, that is endorsed by the management of your organisation?**

Yes

No

Do not know

The answer of this question will be 'yes' when your institution has a formal document that describes the strategy for the **digital preservation** and permanent access to your digital heritage collections.

25. Are your digital collections stored in digital archives that have been set up according to **international standards for **digital preservation**?**

Yes, we have our own digital archive that meets the international criteria for long term preservation

Yes, our digital collections are archived in a publicly managed professional digital archive

Yes, our digital collections are archived in a privately managed professional digital archive

No, we do not have a solution yet for the long term preservation of our digital collections based on international standards

Do not know

Answer this question with 'yes' if your institution is actively involved in safeguarding the digital heritage collections for future generations, based on international standards or best practices.

SECTION 5/7. Digitisation Expenditure

26. Please estimate your annual expenditure on your **digital collections (**total cost of ownership**)**

	Please estimate the budget concerned (€):	Please specify the year concerned:
Institutional expenditure (internal budget):		[drop-down list: 2010, 2011, 2012]
Temporary funded project expenditure (external budget):		[drop-down list: 2010, 2011, 2012]

These budgets should be estimates of the costs related to the initial creation, ongoing maintenance, enhancement and preservation of your digital collections. Please attempt to include the cost of the staff time devoted to digital collection related activities in these estimates. If budget year does not coincide with the calendar, please choose the calendar year that fits best (in terms of the number of months)

Costs can be 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. Examples: selection of materials, acquisition of digital born materials, scanning, descriptive metadata creation, project management.
- Structural costs are the costs needed for the ongoing maintenance, enhancement and preservation of a digital collection. Examples: activities concerning the preservation of digital collections, licences, maintenance of web servers, user outreach and support, management.

27. Please estimate what percentage of the total annual expenditures on creating/acquiring, maintaining, enhancing and preserving your *digital collections* can be assigned to *incidental costs* and what percentage can be assigned to *structural costs*

	Incidental costs:	Structural costs:	
Percentage [should add up to 100%]	... %	... %	100 %

28. Please estimate what percentage of the total annual expenditures on creating/acquiring, maintaining, enhancing and preserving your *digital collections* is spent *In-house* and what percentage is *Outsourced*

	In-house costs:	Outsourced costs:	
Percentage [should add up to 100%]	... %	... %	100 %

The community of libraries, archives and museums would benefit from a better understanding of the costs involved in creating and preserving digital collections. Please help us by providing a more detailed account of your costs.

[] Show table

I. Please estimate what percentages of the *Incidental costs* can be assigned to the following activities

%	Incidental cost category
	Project management
	Selection of material for digitisation
	Acquisition of digital born material
	Logistics (shipment of collection for digitisation, etc.)
	Analogue to 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
	Web design, software development
	Other costs (specify below)
100%	

The sum total adds up to 100%. 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.

Other costs: [input box]

II. Please estimate what percentages of the **Structural costs can be assigned to the following activities**

%	Structural cost category
	Management
	Archiving (storage, including backups)
	Activities concerning the long-term preservation of the digital collection (including research activities but excluding Archiving costs)
	Licences
	Maintenance of web servers and web, mobile and other services
	User outreach and support (including staff time for efforts to promote the use of the digital collections)
	Usage analysis (including user surveys, interviews, and other activities)
	Editorial (including content selection and updating)
	Other costs (specify below)
100%	

The sum total adds up to 100%. 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.

Other costs: [Input box]

29. What is the total number of paid staff (in full-time equivalent) engaged in creating/acquiring, maintaining, enhancing and preserving your **digital collections on an annual basis?**

[Input box]

Include the time of your own institution's staff engaged in activities related to creating/acquiring, maintaining, enhancing and preserving your **digital collections**, including: planning and managing in-house and contracted projects; preparing and digitising materials; enhancing digitised output to widen accessibility.

30. What is the total number of volunteers (in full-time equivalent) engaged in creating/acquiring, maintaining, enhancing and preserving your **digital collections on an annual basis?**

[Input box]

Include the time of your institution's unpaid staff. Volunteers who receive compensation for their expenses (like travel costs) should also be included.

31. From what sources are your digital collection activities funded?

- Internal budgets
- Crowd funding
- National Public grant/subsidy
- Regional/Local Public grant/subsidy
- Private funds and legacies
- Public/private partnership
- Sales of digital items
- Other: [Input box]

Indicate all the sources from which your digitisation activities are funded.

SECTION 6/7. General Notes

32. Please include any information that was not asked for above and that you think is relevant for understanding the nature of activities related to your digital collections.

[Free text field]

Comments on the questionnaire itself can be given in the next question (33).

SECTION 7/7. Questionnaire Evaluation

33. Please include any comments that would help us to improve future issues of this survey.

[Free text field]

O-O-O

Thank you for completing this survey!

More information on the *ENUMERATE* project, and the results it has delivered, is available on www.enumerate.eu.