Grant Agreement 270939

ENUMERATE

Thematic Survey Methodology

Deliverable number	*D2.8*
Dissemination level	*PU*
Delivery date	*March 2013*
Status	*Final after Review Report*
Author	*Gerhard Jan Nauta and Marco de Niet (DEN)*

Revision History

Revision	Date	Author	Organisation	Description
V0.1	2012-10	Gerhard Jan Nauta Marco de Niet	DEN DEN	Draft
V1.0	2012-11	Gerhard Jan Nauta Marco de Niet	DEN DEN	Final
Reviewed by		Natasha Stroeker	Panteia	
		René Vogels	Panteia	
		Jesús Domínguez	DIGIBIS	
		Monika Hagedorn-Saupe	SPK	
		Gordon McKenna	CT	

V2.0	2013-03	Gerhard Jan Nauta Marco de Niet	DEN DEN	Revised (after review)
Reviewed by		Natasha Stroeker	Panteia	
		René Vogels	Panteia	
		Gordon McKenna	CT	

Contents

1 Introduction and overview

The work of ENUMERATE is centred on three surveys, which will be rolled out during the three years of the project. In the first year (early 2012), the focus was on a high level survey (Core Survey 1), as recommended by the Numeric Study Report, on which ENUMERATE builds its activities. In the second year of the project a Thematic Survey will take place, dedicated to the four main topics that ENUMERATE addresses: size of digital heritage collections; costs of digitisation of cultural heritage; use of digital heritage and digital preservation. While the Core Survey is of a quantitative nature, the Thematic Survey will focus on qualitative assessments and methodology development. This is especially for Core Survey 2, because in the third year, the Core Survey will be repeated, in order to create a reliable baseline of intelligence about digitisation of cultural heritage in Europe.

According to the Description of Work the Thematic Survey, as opposed to the two Core Surveys, is:

> "*…more in-depth, (…) designed to assess digitisation, digital preservation, usage and cost in greater detail, in order to build up a more granular picture of overall activity and best practices. The thematic survey will use the original extensive Numeric questionnaire as a starting point, in order to guarantee a high degree of continuity between Numeric and ENUMERATE. However the methodologies for the thematic survey will be based on a thorough evaluation in ENUMERATE expert groups in order to remove the flaws from the Numeric questionnaire, as advised by the SIG-STATS.*"

> [from: ENUMERATE *Annex I to the Grant Agreement (Description of Work) for a Thematic Network*, Version 2.0 (2011-12-2)]

This deliverable describes the overall Thematic Survey schedule, i.e. the interrelated activities, scheduled in time, that are necessary to collect the information as indicated in the Description Of Work. It must ensure that the Thematic Survey rollout is completed effectively. The document is composed of five parts, detailing the steps necessary to do research on each of the four main themes of the ENUMERATE survey project, supplemented with what is necessary to have contextual information in each of the sub-studies. Below we will briefly explain the five parts. After that the parts will be presented in more detail, followed by a tabular representation of the various activities planned.

The main parties involved in running the Thematic Survey are: heritage institutions; National Coordinators; Panteia (the contracted partner); the ENUMERATE Team (in particular the Core Group: CT, DEN, Digibis, SPK); and the ENUMERATE Thematic Network as a whole.

The Thematic Survey will run from November 2012 to February 2013. After that a report summarising the sub-studies will be prepared, the deadline for this is the end of March 2013. Data collected in the Thematic Survey will be published on the ENUMERATE Website. This is scheduled for April 2013. The schedule for earlier activities is based on these deadlines.

1.1 Overall activities (October–November 2012)

The Thematic Survey will not be a single questionnaire distributed among a large group of cultural heritage institutions simultaneously. For each theme and sub-study specific heritage organisations will be selected and approached to collaborate with the ENUMERATE Team and National Coordinators in the various EU member states. To make it easier to collect information, and to build up a reliable database of heritage organisations, in every part of the TS equally structured (contact) information will be collected, preferably with the direct involvement of the National Coordinators. In essence the data collected will be the "scene setting" data that were also collected in the NUMERIC and ENUMERATE Core Surveys (e.g. name of institution, and contact person).

In the Thematic Survey, much will depend on the commitment of the National Coordinators. The ENUMERATE Team will make efforts to create an atmosphere of involvement through consulting the National Coordinators about the themes they think can best be studied in their respective countries and about institutions in their countries that can be invited to participate in the Thematic Survey.

1.2   Theme 1: Size of digital collections (September 2012–February 2013)

The purpose of this part of the Thematic Survey is both to check and review data collected in Core Survey 1 and to further develop the methodology of quantifying the size and growth of digital collections in sufficient detail. This survey module will take on board the recommendations from the Specialist Meeting, held in The Hague on 30 January 2012. The focus will be on improving the problematic "optional table" in Core Survey 1. We will do this by applying three research methods. Approximately 7 countries will be involved in this part of the survey. In each country about 10-15 heritage institutions will be approached that will be asked to compare three ways of measuring the size of collections:

- Measuring collection size and composition based on the records in Collection Registration Systems;

- Detailed measuring in the three most important digital sub-collections of institutions;

- High level, estimate-based measuring, using an improved version of the questions in the ENUMERATE Core Survey.

1.3   Theme 2: Cost of digitisation (October 2012–February 2013)

In this part of the Thematic Survey a method will be developed in order to get more detailed information on the cost of digitisation than could be collected through the questions in Core Survey 1 (refer to question 27 of the CS questionnaire). In the Specialist Meeting devoted to the cost of digitisation (London, 30 March 2012) it was suggested that a workflow approach - Activity Based Costing - should be the basis for the Thematic Survey. The steps consist broadly speaking of designing a workflow on paper; then testing it on a small group of heritage institutions; making improvements to the initial design; followed by a test among a larger sample of institutions.

1.4   Theme 3: Use of digital heritage (September 2012–February 2013)

Here too the purpose is to improve and specify the questions related to accessing digital collections in the Core Survey. After Specialist Meeting 3 in Madrid (6th June 2012) essentially two approaches to monitoring access, use and impact of digital collections stand out. In one approach monitoring access is realised using a more or less extensive questionnaire, to be distributed among a sample of institutions. A drawback of this way of monitoring is that institutions are expected to have the data at hand – which is often not the case - and are willing to fill in an extensive questionnaire. The other approach is doing some sort of action research, i.e. working together with an institution in order to improve internal evaluation procedures. The activities scheduled in the Thematic Survey aim at a combination of both approaches.

In the first part, the questionnaire approach, the ENUMERATE Team will select 30 institutions from a restricted number of EU member states (3 to 5), representative of the different institution types identified in the previous ENUMERATE and NUMERIC surveys. These institutions will be asked to participate in a thorough review of the Core Survey and other access related questions (based on work done by other surveys), with special attention to the usefulness of the questions and the feasibility of answering them.

In the action research approach either the same institutions or another sample (to be determined) will be asked to monitor a particular stream of digital collection information, viz. web traffic. Institutions will be asked to install web analytics software (probably Google Analytics) and monitor during a fixed period of time a small number of standard indicators. The museum sector has seen web analytics initiatives in recent years. ENUMERATE will especially do research on cross-sectoral monitoring. In this part of the Thematic Survey ENUMERATE seeks collaboration with partners in existing web analytics projects, such as Museum Analytics or Culture24.

1.5   Theme 4: Digital Preservation (September 2012–February 2013)

Monitoring digital preservation policies and practice is an activity that is still relatively new to the cultural heritage field. Therefore the ENUMERATE Team has decided, in consultation with a group of specialists that met in The Hague on 5 October 2012 to focus on the development of a small but balanced set of survey questions. The approach will be a combination of desk research, the consultation of a focus group on the topic, and a synthesis of outcomes. The ENUMERATE Team has experienced that the need to monitor digital preservation in the domain of cultural heritage is emerging in several places. Realising synergy with other plans and projects will be part of the activities. The main activity, the focus group session, is scheduled in the month of January 2013.

2 Role of the National Coordinators

2.1 Involvement of the National Coordinators

As stated above, the commitment of the National Coordinators is crucial to tap into data and knowledge available at cultural heritage institutions across Europe. For each of the themes in the Thematic Survey the ENUMERATE Team will discuss options for participation with a selection of the National Coordinators. In this way the workload for the National Coordinators can be shared equally. National Coordinators are invited to volunteer for a specific theme. Their main activities will be to identify relevant institutions, to provide contact details and, if they desire to do so, be involved in evaluation of the results from their country.

In this phase we will also ask the national coordinators about the state of the national strategy on digital preservation in their country (**Which countries have an established national preservation strategy?)

- Is there an established national preservation strategy in your country?
- What is the status of this strategy?
- And is there an established national digital preservation infrastructure in your country?
- What is the status of this infrastructure?

This was a result of the 5th October meeting in Den Haag.

2.2 Database of contact details (sampling database) design and construction

From the data collected in the ENUMERATE Core Survey an initial database of contacts will be constructed by Panteia in consultation with DEN and Digibis. The design of the database will be fairly simple. It will contain contact information about the institution, their role in the ENUMERATE network so far and their role in the Thematic Survey. This database is then presented to the National Coordinators, who are expected to select institutions or submit suggestions for new institutions to participate in the Thematic Survey. Institutions that did not yet participate in the ENUMERATE Core Survey will be added to the database. This is an on-going activity, up to the end of December 2012.

2.3 Schedule of activities

Contact database design and construction is an ongoing process during the start-up phase. It will be finished by the end of December 2012.

3 Theme 1: Size and Growth of digital collections

3.1 Methodology development

Acquiring in-depth information on the size and growth of digital collections is a real challenge for any monitoring initiative in the domain of cultural heritage. The varying size of memory institutions - with few or many staff involved in collection management and digitisation - and the circumstance that overall collection management information is often incomplete (the precise size of analogue and digital collections is often not known), causes ENUMERATE to choose a multifaceted approach. In a restricted sample of institutions quantifying digital collections is done in three slightly different ways. Each of the selected institutions will participate in all three approaches to quantifying the size of digital collections. Purpose of this approach is to assess the variance that results from using different methods of counting. The institutions in this part of the TS will also be asked to evaluate the practicability of the different methods, which may perhaps lead to varying conclusions about the appropriateness of different methods in different sub-domains of the heritage domain. Below these methods are explained in more detail.

Theme 1a: Measuring based on the records in Collection Registration Systems

In this part of Theme 1 no effort will be taken to collect data other than what is registered in collections registration systems (CRS). This can relate to both physical and digital objects. For this part of the survey staff members with a responsibility for information management at the level of collections will be addressed. An important step will be to determine the number of separate collection registration systems in use. The number of records per system must be established. Respondents will also be asked to supply the list of object types in use and if possible the number of records per object type. This approach will contribute to acquire a better understanding of and better definitions for measuring cultural heritage collections.

Theme 1b: Detailed measuring in the three most important digital sub-collections

In addition to this global way of measuring, the TS will invite institutions to measure the three most important digital collections in their holdings. As was suggested by the Specialist Meeting, it is worthwhile to try and approach the measuring from the digital perspective instead of the physical perspective. This means we will use a similar table as the optional question #14 of the ENUMERATE Core Survey 1, but this time with a (draft version of a) classification of digital collections. Starting point in this approach is a list of object types from a report published by the DEN Foundation in 2010).[1] The initial list will be complemented with the object types taken from the ENUMERATE Core Survey.

Methodology development will consist of improving this classification, deciding on the final units to measure and improving definitions of digital collections. For this part of the survey staff members with a responsibility for (digital) collection curation will be addressed, i.e. those who have in depth knowledge of the (composition) of the specific collections.

Theme 1c: High level, estimate-based measuring, using an improved version of the questions in the ENUMERATE Core Survey

The third part of the survey is a slightly extended version of the present Core Survey 1 questions about estimated percentages of digitised and to be digitised collections. In the questions where the respondent is asked to indicate the object types that are include in the estimates an additional qualifier - estimated percentages of the relative volume of these collections - will be asked. These questions will be sent to managers of the institutions, who should have generic knowledge about the collections.

With these three approaches, partly chosen to deal with complications inherent to the different sizes of cultural heritage institutions, ENUMERATE uses a layered approach to collecting data. This approach was inspired by the recent ITHAKA S+R Survey on *Sustaining Digitized Special Collections*. ITHAKA S+R designed three separate questionnaires in order to collect data on the level of: 1. The institutional perspective; 2. Collections in the aggregate; 3. Single collections. The ITHAKA methodology provided a mechanism for passing the request to participate in the survey from a high (management) level in the organisation to more specialised levels, where there is probably a better knowledge of collection specifics.

Methodology development for Theme 1 runs from September to November 2012.

---

[1] 'Born-digital heritage materials at selected Dutch heritage institutions', technical report, The DEN Foundation (The Hague, 2010). Refer to http://www.den.nl/art/uploads/files/DEN/BorndigitalHeritagematerials_Netherlands_ENG.pdf (link accessed 20/11/2012).

3.2 Desk research and check-up of Core Survey 1 data on the size of digital collections

In this phase of the survey the outcomes of the size and growth estimations from Core Survey 1 will be reviewed by a modest number of specialists in the various institution types. This review will contribute to improvement of the classification of physical collections and their measuring units.

3.3 Guidance materials

As in the earlier Core Survey there will be both materials aimed at guiding the National Coordinators in working together with the institutions, and materials guiding respondents when providing information about their institutions. In principle questionnaires for Theme 1 will be made self-explanatory. For most of the Thematic Survey activities guidance materials will therefore be kept as simple as possible. The preparation of guidance materials and questions on a meta-level (questions about the actual work being done) is a task for DEN in collaboration with Panteia.

For Theme 1 guidance materials can in part be adapted from the Core Survey guidance materials. Questionnaire guidance will be limited to introductory remarks (covering letter) and an improved version of terms and concepts in the glossary. If feasible the tables where institutions can indicate the size of their collections and the degree to which they are (being) digitised, will be made expandable. New will be a table with a classification of (born) digital heritage collections. This will be based on work done by DEN in a previous national project, on the classification tools collected in the first year of the ENUMERATE project and on the input from the Specialist Meeting on this topic.

3.4 Translations

All survey materials will initially be developed in English. National Coordinators may choose for a translation of the questionnaire and guidance documents so that the target institutions selected can more conveniently understand it. To ease the work to be done and to reduce the workload for individual participants translations the ENUMERATE consortium partners are asked to translate the survey materials into the main European languages represented by the partners (English, French, German, Spanish). Digibis will be asked to support this translation process with their online tool used for the Core Survey 1.

3.5 Selecting institutions

Selecting institutions is an activity done in a collaboration of the ENUMERATE Consortium Partners and the National Coordinators. For Theme 1 about 7 EU member states will be involved. The National Coordinators in these countries are expected to select about 10 to 15 institutions each: n=105 (max.). In the Thematic Survey methodology document details will be given concerning the precise method of selecting the institutions.

3.6 Online survey construction

All three approaches to survey this theme will result in online questionnaires, to be developed by Panteia, in close cooperation with DEN, Digibis and the National Coordinators involved. This is done in order to stay close to later versions. For translation purposes, an offline version of the questionnaire - e.g. in the form of a spread sheet - can be considered. In order to improve feedback some questions on a meta-level will be added to the online questionnaire. The online questionnaires will be mounted on a server hosted by Panteia, and it will be made accessible through the ENUMERATE website.

Online survey construction will be done in the month December 2012. The deadline for the publication of the online questionnaire is 4 January 2013.

3.7 Online survey implementation

On behalf of the consortium, Panteia will compile a list of institutions that will be invited to participate. The actual survey will run from 4 to 31 January 2013. Coordination is a task for Panteia, in cooperation with DEN and DIGIBIS. Three tasks stand out here: maintaining the database of invited institutions, providing support to the National Coordinators, and monitoring and securing responses. Since institutions are asked to share information on quantities of materials in their analogue and digital collections there is a fair risk of faults in the data collected. Therefore in the implementation phase monitoring the incoming data will be an on-going activity and actions to correct where necessary or to ask respondents to clarify responses will be taken ad hoc. A more thorough cleaning and harmonising job is scheduled after the survey period.

3.8 Urging non-responders to respond

During the four weeks the survey will be online, the consortium contacts (National Coordinators) will be available to answer questions about the survey and if necessary urge non-responders to take action. This is a shared responsibility of the National Coordinators and the ENUMERATE Team.

3.9 Cleaning and harmonising returned questionnaires

Any corrupted or ambiguous data in the survey responses must be corrected or left out. This part of the work is scheduled between 28 January and 8 February 2013.

3.10 Data analysis

Data analysis and preparing the interim report is scheduled between 4 February and 1 March 2013.

3.11 Schedule for Theme 1

Theme 1 will run from September 1st 2012 to February 8th 2013. Data analysis is scheduled in the month of February.

4 Theme 2: The cost of digital collections

4.1 Activity based costing methodology development

The schedule for Theme 2 consists of two separate activities, not overlapping in time. First a workflow model for calculating costs must be designed, or rather, adapted from existing examples. DEN will design the model, based on earlier and present experience with this kind of research.[2]

The model will be tested in about 15 test cases. After that the model is refined on the basis of an evaluation of test results. The next step is using the improved cost model in a larger sample (n=90) of memory institutions.

4.2 Selecting institutions

Again selecting institutions is done in close collaboration between the ENUMERATE Team with the National Coordinators. To realise the test case part of the survey process the ENUMERATE Team needs about 15 institutions that have some expertise in cost calculations or that are in the process of doing digitisation projects in which the monitoring of costs is transparent or especially needed. Three National Coordinators in countries that are in the frontline of digitising cultural heritage will be asked to assist in tracing these institutions (target: 5 institutions/country).

In the second part of Theme 2 about 7 EU member states will be involved. The National Coordinators in these countries are expected to select about 10 to 15 institutions each: n=105 max. In the Thematic Survey methodology document details will be given concerning the precise method of selecting the institutions.

Selecting institutions can be done fairly early in the survey period; the deadline is 16 November 2012.

4.3 Test case phase workflow survey

Workflow test cases will be done between 19 November and 14 December 2012. Prior to that the survey instruments will have to be developed, but since the number of participating institutions in this phase is small, the construction of the instruments will be parallel to methodology development.

4.4 Workflow test case reviewing and synthesis

The outcomes of the test case phase will be important in a review of the initial workflow design. A new improved workflow model will be developed by DEN in the month of December (3 December 2012-4 January 2013).

4.5 Workflow survey implementation

The second part of the workflow survey is planned during the month of January 2013 (7 January to 1 February 2013).

Panteia will take care of constructing the workflow questionnaire for part two of the survey on cost. Whether the questionnaire will be digital (online or offline) or on paper is still open to discussion (refer to Deliverable 2.8, the methodology document), but since the final aim is to have a tool that is both focussing on the most important costs in creating digital collections and easy to handle, the questionnaire must be simple and the actual implementation mode will not have an influence on the planning cycle.

4.6 Cleaning and analysing returned questionnaires

Any corrupted or ambiguous data in the survey responses must be corrected or left out. This part of the work is scheduled between 28 January and 8 February 2013.

4.7 Data analysis

Data analysis and preparing the interim report is scheduled between 4 February and 1 March 2013.

4.8 Schedule (blueprint) for Theme 2

Theme 2 methodology development is scheduled in the months October and November 2012. The test cases should be finished by mid December, whereas the actual workflow survey will be done in the month of January 2012. Data analysis is scheduled in the February 2013.

---

[2] 'Rekenmodel digitaliseringskosten (Rekenmodel)', techbical report, The DEN Foundation (The Hague, 2010). Refer to: http://www.den.nl/standaard/202/Rekenmodel-digitaliseringskosten (in Dutch).

5 Theme 3: the Access and Use of digital collections

5.1 Methodology development

As was described under paragraph 1.4 the subject of measuring access and use (and impact) of digital collections will have a dual approach: one part will be more or less a conventional survey (the 'questionnaire version'), specifically aimed at improving and deepening the questions that were posed in Core Survey 1. The methodology for this part of the survey should be finished by the end of November 2012. The three sources for this part of the survey are:

- The questions from the ENUMERATE Core Survey, adapted on the basis of the various comments and that were collected during and after the implementation of the survey, early 2012;

- The questions from the more extended NUMERIC questionnaire;

- Questions used in another survey done in the Netherlands [3]

The second part (the 'data analytics version') is with a proviso: it can only be designed in close collaboration with one or more parties already active in the field of quantifying the way institutional websites are being used, and the activity/presence of memory institutions on social media like Facebook, Twitter, and Flickr. Since it is difficult to predict, because of the interdependencies, whether this is achievable in the time span set for the Thematic Survey as a whole - much will depend on the interim appointments that can be made - we have planned a *Data Analytics feasibility* study, possibly with a small action research project in the months December and January. It is impossible to plan in more detail now, therefore this part is of the schedule is relatively open.

5.2 Desk research and updating an overview of monitoring initiatives

In preparation of Core Survey 1 an inventory of surveys and different ways of monitoring evaluating and measuring the value, use and impact of digital collections was produced by the ENUMERATE Team. Because recent years have witnessed a lot of activity in this area - undoubtedly induced by the increasing importance attached to business models and accountability in the heritage domain -, with many stakeholders that will be interested in a thorough overview, the ENUMERATE Team will reconsider the way information is collected and published. Especially making it a social responsibility of the heritage domain as a whole to update the overview is the goal that the ENUMERATE Team will strive to achieve.

Updating the existing inventory is an explicit task in the months September and October 2012. After that the social bookmarking mechanism should be in place.

5.3 Guidance materials

Where the sample in this part of the Thematic Survey will be relatively small, guidance materials will be limited and since we plan for a close cooperation with about 3 National Coordinators, guidance materials will only be needed in the phase of collecting information in the institutions. Ideally questionnaires of interview protocols for Theme 3 will be made self-explanatory.

Guidance materials can partly be adapted from the Core Survey guidance materials. Questionnaire guidance will be limited to introductory remarks (covering letter) and an improved version of terms and concepts in the glossary.

5.4 Translating the questionnaire

Depending on the choice of the 3 to 5 countries where Theme 3 will be surveyed, some translation work may be necessary. If possible all work on Theme 3 will be done in English, but the National Coordinators will have a say in the desirability of producing a translation. Special care should be given to the accuracy of that translation. If a control mechanism cannot be implemented, the ENUMERATE Team should waive translation efforts.

5.5 Selecting institutions

In the questionnaire version the ENUMERATE Team will select 30 institutions from a restricted number of EU member states (3 to 5), representative of the different institution types identified in

---

[3] Henk Voorbij, 'The use of web statistics in cultural heritage institutions', *Performance Measurement and Metrics*, jaargang 11 (2010) nummer 3, pp. 266-279. See also: http://www.den.nl/art/uploads/files/Publicaties/Webstats_summary_ENG.pdf (link accessed 20/11/2012).

earlier [ENUMERATE and NUMERIC] surveys. These institutions will be asked to participate in a thorough review of Core Survey and other access related questions [from other surveys], with special attention to the usefulness of the questions and the feasibility of answering them.

5.6 Online survey construction

Preferably for Theme 3 again an online questionnaire will be used - advantages are in the reusability of questions and guidance materials - but when it is considered more practical to work with other survey instruments here [- the number of institutions will be small -] questionnaires on paper etc. may have to be developed.

Survey/questionnaire construction is a task for Panteia, possibly in collaboration with DIGIBIS.

5.7 Survey implementation

Where the survey sample is modest the survey will be organised in a relatively short time span: from the end of December 2012 (24/12/2012) to 11 January 2013, with an option to extend the survey by one week.

5.8 Urging non-responders to respond

This activity is scheduled in the week of 7 to 11 January 2013.

5.9 Cleaning and harmonising returned questionnaires

After having collected responses the data must be cleaned and harmonised. This activity is scheduled in the week of 7 to 18 January 2013.

5.10 Data analysis

Panteia will collect and collate the data as submitted by the institutions that are going to participate in this part of the Thematic Survey. Analysing the outcomes of the workflow survey is a task for Panteia too, but close cooperation with DEN is necessary, since domain knowledge may be crucial to interpreting the results. If necessary, feedback from the National Coordinators will be asked.

After collecting and analysing the survey outcomes, a short report will be prepared by Panteia, summarising main outcomes. It will be reviewed by the ENUMERATE Core Group. The report will be discussed by e-mail with other stakeholders (Advisory Board, ENUMERATE Consortium).

5.11 *Data Analytics* feasibility study and action research

In the action research version a small sample of institutions will be asked to monitor web traffic, using Google analytics, or some other web statistics program (to be determined). Institutions will be invited to install and monitor a small number of standard indicators over a fixed period of time. This part of the Thematic Survey is dependent on collaboration with partners in existing web analytics projects [Museum Analytics? Culture24?].

It is difficult to plan at this stage when an analytics experiment can be done. The provisional time schedule is January-February 2013. We have not included this in the blueprint below, all the more since it is probable that others than the ENUMERATE Team might take the lead here.

5.12 Processing *Data Analytics* research outcomes

Data analysis and preparing an interim report for this part of the Thematic Survey cannot be planned yet.

5.13 Schedule (blueprint) for Theme 3

The actual survey is scheduled early in January 2013. The data analysis phase has a deadline on February 15th. For the optional data analytics phase a longer time span should be reserved. Depending on the outcome of negotiations with interested co-organisers this part of the Thematic Survey may run until the end of February.

6 Theme 4: The preservation of digital collections

As was indicated in the ENUMERATE Description of Work (version December 2011) there is no statistically valid basis upon which to benchmark digital preservation, or through which to normalise or harmonise issues such as cost and format in the context of digital preservation. In spite of considerable progress in raising the level of awareness, especially among governments and the more advanced and often larger heritage institutions, and in spite of the development of technically advanced models, standards and practical solutions, there have been disappointingly few initiatives to improve quantification methods. Specialists have therefore welcomed the idea of developing a framework for monitoring digital preservation policies, plans and activities in digital collection holding institutions. Because of the relative immaturity of the area - it was estimated that the bulk of the memory institutions is not prepared to answer detailed questions on digital preservation issues - ENUMERATE has decided to follow a tiered model of information gathering, where in the initial phase much time and effort will be spent on making explicit the inside knowledge and opinions of experts. A focus group seems to be perfectly suitable to assess the options for measuring digital preservation practice. See Appendix 8.6 - *Monitoring and Measuring Digital Preservation Practices* (Brainstorm session, The Hague October 5, 2012 - for further details.

6.1 Desk research

The usual desk research will take place especially during the first months of the survey period, but it will be setup as a continuous community effort. References will be collected using an online social citation and bookmarking tool. (Presently Delicious and CiteUlike are in use, but the ENUMERATE Team plans to switch to Diigo.) Also documentation suggested on the 5th October meeting will be used.

6.2 Initial selection of Digital Preservation survey topics

From the desk research an overview of topics addressed in earlier surveys will be compiled. During the Brainstorm session (see below) the usefulness of addressing such topics will be assessed. (Is it possible to answer these questions? Are these questions suitable in a focus group? Etc.)

6.3 Brainstorm session: the monitoring of Digital Preservation status

An informal meeting on monitoring and measuring digital preservation practices in the EU will explore the options of organising a focus group early in January 2013. The event has taken place at the Den Foundation, in the Royal Library (KB) building in The Hague, Friday October 5th, at 10:00 am. Invited for this session was a small group of local (mostly Dutch) specialists.

As a preparation to the meeting all invited were asked to reflect on the type of intelligence that would be needed, both in order to support individual institutions in the management of their collections - an essential incentive to participate in any kind of survey effort - and also, in the long run, to support national and EU governments in adopting and implementing digital preservation policies (see above).

A second request to the brainstorm group was to review a few questions on the topic of digital preservation that were part of the ENUMERATE Core Survey (February-March 2012).

Main results from the 5th October meeting:

1) We will test the standards for digital preservation in the Thematic Survey:

- Do you know / are you familiar with PREMIS?
- Do you have your data available in PREMIS?
- In which way does this help your organisation? (think of a better formulation)

Ask for all four standards: OAIS (see above), PREMIS, DRAMBORA, and TRAC

2) Input for the focus group in January:

- How to reach the target groups;
- Results of surveys done in this area;
- Harmonisation of planned surveys in this area;
- Test of questions;
- Checks on definitions.

6.4 Formation of Focus Group on Digital Preservation monitoring

An initial list of experts to be invited resulted from the brainstorm session. Early in November the list of experts will be finalised and the focus group will be announced.

6.5 Focus Group preparation

Based on desk research, the results of the brainstorm group, and a consultation of key figures in the field of digital preservation and curation an agenda for the Focus group will be prepared in the second half of December and the first week of January.

6.6 Focus Group event

The actual focus group will be organised taking into account the agenda of the international digital preservation community. A possible event to align with is the *8th International Digital Curation Conference*, which will take place in Amsterdam, from 14 - 17 January 2013.

6.7 Focus Group report

Following the Focus Group event a concise report, presenting the principal outcomes of the meeting will have to be prepared. The contents of the report will be assimilated in the final Thematic Survey report. (See below.)

6.8 Schedule (blueprint) for Theme 4

Preparation of the focus group meeting will be done in the months September to December 2012. The actual focus group meeting will be around January 17th 2013.

7 Final report Thematic Survey and publication of survey data

7.1 Thematic Survey report

The Thematic Survey report must be a compiled on the basis of the four sub reports discussed above. It is the second report in a series of ENUMERATE publications: Statistical Monitoring of Digitization of Cultural Heritage in Europe, and it will be published on the ENUMERATE website. The report presents the results of the Thematic Survey, and will contain where possible references to Numeric Final Report, in order to show developments since the Numeric project. The report must be announced through all digital channels that ENUMERATE can employ.

7.2 The ENUMERATE Data Platform

The data collected in the Thematic Survey events will be anonymised and published on the ENUMERATE Data Platform. Depending on the available project budget, tools to assess and report on the statistical data can be part of the services offered at the Data Platform. The ENUMERATE Team will make efforts to search for additional funding to support the development of a sustainable solution.

The technical design and implementation of the Data Platform is an activity of Digibis, whereas preparing the content to be published is a task for Panteia, in collaboration with Work Package 3 leader Digibis. This will take place in the month of April 2013. The actual launch of the platform is foreseen no earlier than the summer of 2013.

8 ANNEXES

8.1 ENUMERATE THEMATIC SURVEY TIME SCHEDULE
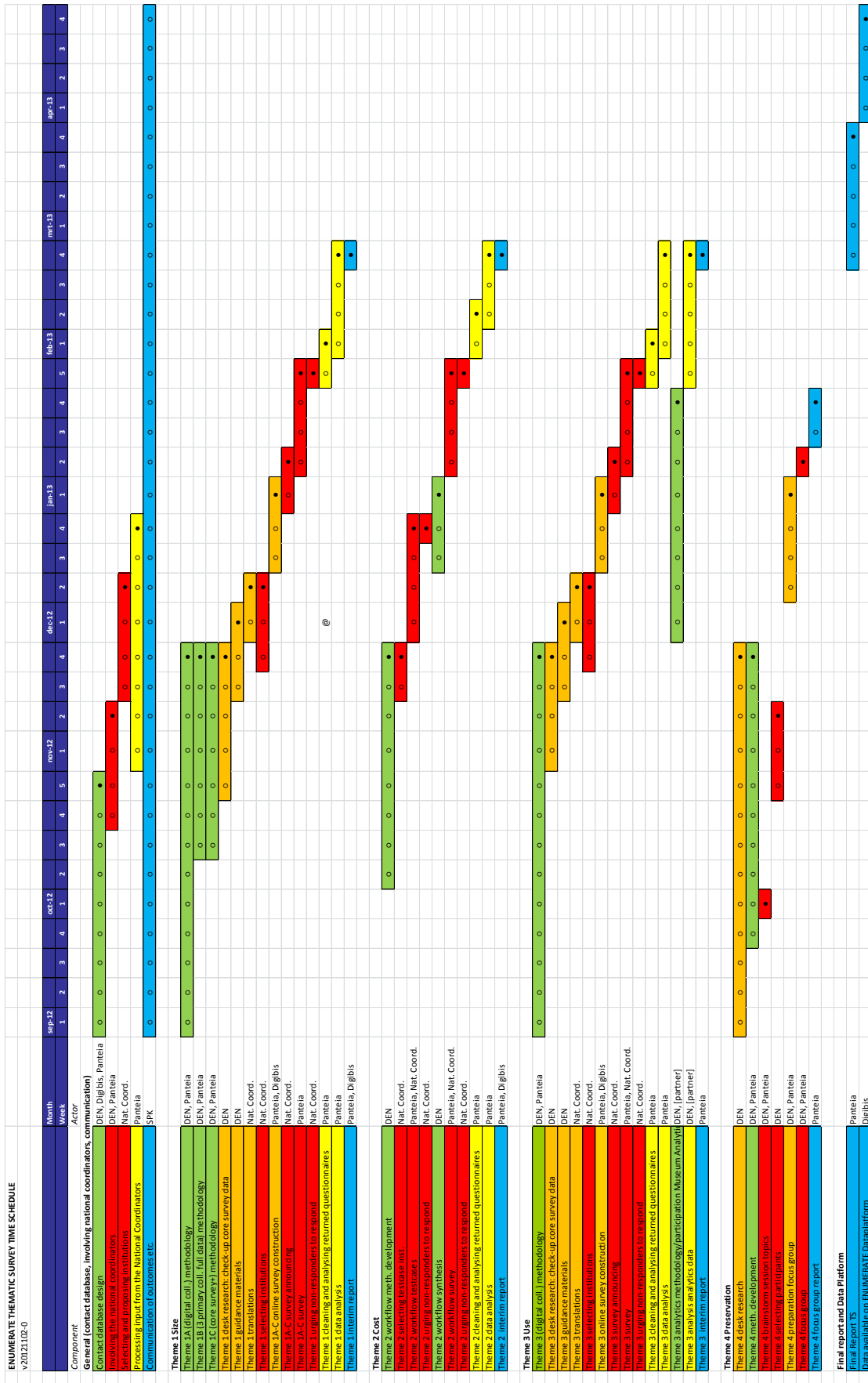
8.2   MEMO Involvement of the National Coordinators in Thematic Survey activities

MEMO
From   *ENUMERATE Team (DEN & Panteia)*

To   *National Coordinators*

Date   *November 13th, 2012*

Subject *Involvement of the National Coordinators in Thematic Survey activities*

ENUMERATE Project

The work of ENUMERATE is centred on three surveys during the three years of the project. In the first year (early 2012), the focus was on a high level survey (Core Survey 1). In the second year (November 2012 to April 2013) a Thematic Survey is scheduled. It will be dedicated to the four main topics that ENUMERATE addresses:

1.   Size of digital heritage collections;
2.   Cost of digital heritage collections;
3.   Use of digital heritage collections;
4.   Digital preservation.

Thematic Survey

In the Thematic Survey, each topic will get its own sub-study. While the Core Survey is of a quantitative nature, the Thematic Survey will focus on qualitative assessments and methodology development, especially related to Core Survey 2, since in the third year, the Core Survey will be repeated, in order to create a reliable baseline of intelligence about digitisation of cultural heritage in Europe. The number of institutions will be considerably smaller in the Thematic Survey than was the case in Core Survey 1. See the appendix (below) for more information on the Thematic Survey.

> ***For the Thematic Survey to become a success we need the involvement and participation of the National Coordinators!***

What do expect from the National Coordinators?

We would like each country to be involved in at least one of the sub-studies of the Thematic Survey, but when national programs or policies make it desirable to participate in more than one sub-study this is of course possible. Together with the National Coordinators of the countries we will select cultural heritage institutions to include in the particular part of the Thematic Survey.

If you take part in the Thematic Survey, in which part do you want to be involved?

<< you can select more than one sub-study, although initially you will be invited to participate in only one of the options >>

☐   1. Size of digital heritage collections

☐   2. Costs of digital heritage collections (A)

☐   2. Costs of digital heritage collections (B)

☐   3. Use of digital heritage collections (A)

☐   3. Use of digital heritage collections (B)

Other questions we hope you can answer

Question 1. National statistics offices

For building c.q. constructing a database of institutions for each country we would like to know with which national statistics offices each country has warm contacts. So, do you have contacts with statistics offices and if yes, with which national statistics offices?

☐ YES, with ………………………………………………..……….. *<< write the names down here>>*

☐ NO

Question 2. National strategies

We are interested to know whether there is a national digitisation strategy and/or a national digital preservation strategy and/or a national digital preservation infrastructure in your country. Could you please indicate which of these does exist in your country?

	YES	NO
National digitisation strategy	☐	☐
National digital preservation strategy	☐	☐
National digital preservation infrastructure	☐	☐

Question 3. The need for translations

Do you think it is sufficient for all questionnaires of the Thematic Survey to be in English or do you think other languages are really needed? If the latter is the case, are you prepared to support us in making translations?

☐ YES, Thematic Survey in English is sufficient

☐ NO, a translation is needed, namely in ………………………….. *<< please mention language >>*

☐ YES, I am prepared to support the translation

☐ NO, I am NOT prepared to support the translation

Question 4. Any other remarks

Please let us know if there is anything else you think is important to consider before initiating the Thematic Survey:

……

……

……

Please send us your answers as soon as possible, but before the beginning of December 1st 2012 at the latest.

When we have received all responses of the National Coordinators we will make a selection of countries and institutions we will involve in the various parts of the Thematic Survey. This will be done in close consultation with the National Coordinators.

See the Appendix below for further details.

Appendix: The Thematic Survey in Overview

Sub-study 1	*Size of digital heritage collections (7 countries needed)*
Subject	Improving the problematic 'optional table' in Core Survey 1
Method	Comparing three ways of measuring the size of collections by means of a survey (online)
Needed	7 countries to provide 10-15 heritage institutions
Period	January 2013

Sub-study 2	*Costs of digital heritage collections (10 countries needed)*
Subject	Testing the workflow approach (Activity Based Costing) in a small group of cultural heritage institutions (A), followed by a test among a larger sample of institutions (B)
Method	A. Qualitative in depth test on a small group of institutions (personal involvement, by email, telephone, etc.) B. Larger test on a larger group of institutions (online survey)
Needed	A. 3 countries, providing 5 institutions in their country (n=15) B. 7 countries, providing 10-15 institutions in their country (n=105)
Period	A. December 2012 B. January 2013

Sub-study 3	*Use of digital heritage collections (5 countries needed)*
Subject	Improving and specifying the questions related to accessing digital collections in Core Survey 1
Method	A. Thorough a review of the Core Survey 1 on usefulness (online survey) B. Monitor a particular stream of digital collection information (by means of web analytics software) in close collaboration with other initiatives in the field)
Needed	A. 3 to 5 countries, 30 institutions in total (selected from the ENUMERATE and NUMERIC surveys B. the same 30 institutions
Period	A. January 2013 B. December 2012 - January 2013

Sub-study 4	*Digital preservation (no countries needed, experts / specialists)*
Subject	Consultation with specialists, since this is a relatively new topic
Method	Focus Group
Needed	No involvement of National Coordinators needed
Period	January 2013

8.3 *Improving the methodology for measuring digital heritage collections (Report on Specialist Meeting 1, The Hague, January 27, 2012)*

This document is the result of the first of four Specialist Meetings of the ENUMERATE project, which was held at 27 January 2012 in The Hague.

ENUMERATE Specialist Meeting 1

Date: **27th January 2012**

Venue: **Digitaal Erfgoed Nederland (DEN), The Hague, The Netherlands**

Participants:

1. Europeana: Annelies van Nispen
2. IFLA: Ingeborg Verheul
3. ICOMOS / EU-CHIC: Rand Eppich
4. APEX: Gerrit de Bruin
5. EGMUS: Monika Hagendorn-Saupe
6. Rijksmuseum Amsterdam: Lizzy Jongma
7. FIAT / Sound & Vision: Annemieke de Jong
8. CERL: Ivan Boserup
9. Special Collections: Chantal Keijsper
10. ENUMERATE: Sjoerd Bakker
11. ENUMERATE: Gerhard Jan Nauta
12. ENUMERATE: Marco de Niet

Introduction

This document has been created by the ENUMERATE Project, an EC-funded project, led by Collections Trust in the UK. The primary objective of ENUMERATE is to create a reliable baseline of statistical data about digitization, digital preservation and online access to cultural heritage in Europe.

The ENUMERATE Project will conduct three surveys over the course of 2012 and 2013. These surveys are aimed at the following topics:

  • Growth of Digital Collections (supply)

  • Usage of Digital material (demand)

  • Costs of Digitisation (economics)

  • Digital preservation practices (sustainability)

ENUMERATE builds on the results of the NUMERIC project (2007-2009). This was a ground breaking initiative to create a framework for the gathering of statistical data on digital cultural heritage. ENUMERATE will improve and refine the methodology from NUMERIC and will bring the data online for re-use.

In order to improve upon the methodology of the NUMERIC project four specialist meetings are scheduled to take place in the first half of 2012. In each of these meetings one of the topics of the ENUMERATE project will be discussed. The outcome of these meetings will contribute to the creation and improvement of tools that are part of the statistical framework of ENUMERATE.

The meeting was dedicated to the methodology to measure the size and growth of digital heritage collections in relation to the physical collections. The goal of the meeting was to discuss the need for such type of research, the usefulness of a common methodology and the difficulties encountered so far in ENUMERATE and NUMERIC. The Specialist Meeting also took on board the recommendations from the SIG-STATS (EU Special Interest Group on Cultural Heritage Digitisation Statistics).

During the meeting three main issues were discussed:

a. Usefulness of measuring size and growth of digital heritage collections

What are the main motivations for memory institutions to address the issue of size of their collections and growth of their digitisation activities? And why would it be useful to compare them among themselves and across domains?

b. Creating a generic typology of heritage collections

Secondly, is it desirable to develop a generic typology of heritage collections and object types to be used in statistical surveys? In the short term, such a typology would ease the collecting and comparison of data and, in the long term, a generic typology may be of great use to individual institutions (e.g. harmonisation in collection description systems), future research and policy makers.

c. How to deal with heritage collections that cannot be counted easily

Finally, a set of problematic classifications that were part of Numeric were yet to be elucidated. These classifications (Born digital, newspapers, archival records and monuments) are either too broad, vague or otherwise unclear. This may be a matter of definition, choosing the right unit to measure or otherwise.

The meeting was concluded with a round to get suggestions for the next ENUMERATE (Thematic) Survey.

Usefulness of measuring size and growth of digital heritage collections

Before the usefulness of measuring size and growth was discussed those present thought it necessary to reflect on the varying interpretations of the word "digital heritage collections". A few issues did arise:

1. A common understanding of what is cultural heritage is necessary: in the Mediterranean area the tendency is to use the concept primarily for monuments and sites. In Northern European countries in particular museums and archives are associated with it. Common ground is in the idea that objects of cultural heritage are man-made things (as opposed to natural heritage), worthwhile preserving for future use. Curatorial care is part of the mission of cultural heritage institutions.

2. What is meant by digitisation? Collections of digitised cultural heritage objects are not necessarily "digital heritage collections", i.e. collections that are being built in order to safeguard the objects for future use, while offering access for present day users. A lot of digitisation is not meant for access. For example: in the domain of monuments and sites heritage objects are being digitised in order to make work easier for those preserving or repurposing the physical objects. Such digitised resources are not necessarily "digital heritage collections". They are tools in the service of something else.

3. But the ideas about what is, and what is not, cultural heritage is subject to change. Prints were once seen as mere reproductions, they are considered cultural heritage objects now. Books can evolve from mere vehicles of scientific research and scholarly debate, into objects of cultural heritage, expressing the convictions and beliefs of a certain era. Twitter messages are now often valued as worthless, but who knows how future generations will perceive tweets? As a consequence there may be diverging opinions about what to include or exclude in these measurement operations. It is part of the cultural heritage profession to come up with new /adapted selection criteria.

4. Quality needs to be addressed here too. It is not useful to measure size and growth if nothing is known about the quality standards applied.

5. Monitoring size and growth of digital heritage collections as such is only part of the story. As one of the participants said: "It is important to know what you have, but it is more important to know what you can (afford to) keep in the long run." In other words: measuring the actual status of digitisation should be related to understanding and managing the total costs of ownership, which includes digitisation costs, but also the costs of digital preservation.

6. Speaking of the costs of digitisation and long-term sustainability of digital heritage collections, some felt that the target groups in digitisation efforts should somehow come into focus: "What costs you want to spend on digitisation is dependent on your goals. Access may be an important goal. New research questions are an important goal for us [a research library] too, not [even] preservation in the first place." Clearly the aims and missions of the organisations digitising cultural heritage may vary a lot.

All in all the question of the usefulness of measuring size and growth of digital heritage collections was met with a resounding "Yes." It was considered useful for various reasons:

a. Efficiency of business

Information about the size and growth of digital collections is valuable for an institution's decision-making process. Plans based on a reliable baseline of data can be carried out more efficiently and, as a result, cost effectively.

b. Accountability for investments done

Digitisation efforts are often financed with the aid of public money, grants or external investments. Reliable information about the resulting growth of digital collections is an invaluable instrument when reporting to these financiers.

c. Projecting future investments

Digital collections cost money to create, but also to sustain. Information about the size of digital collections can be used to create reliable estimates of the costs of long-term sustainability. This has implications for the selection process as well. If there is a limited budget choices must be made, but: "In order to choose you have to know: how much material is there? How much needs to be digitised? And this is not known yet."

d. Understanding the transformation of digital information into digital cultural heritage

The ideas about what constitutes cultural heritage change over time. We are still in the early stages of the emergence of digital heritage. It is unclear which born digital materials will be considered as cultural heritage in the future. In that respect it is worthwhile to note that in the digital world, everything is information. In the physical world, the distinction between an object and information about an object is clear. In the digital world, where complex digital objects emerge, the distinction gets blurred.

e. Collaboration

Harmonised data about the size and growth of digital collections is essential to efforts like Europeana, where the ambition is to digitize cultural heritage on a large scale. If that is the ambition, it is useful to know how much has been digitised: "these figures/numbers of digital and digitized heritage materials are very important for Europeana."

One of the participants stated: "It is very important to have a mutual vocabulary." And the aspiration is not new. In earlier research/surveys measuring the size and growth of collections was done by way of meticulously counting actual numbers of analogue objects (paintings, prints, books, photographs, three-dimensional objects, etc.) and relating these to the digitized outputs on the basis of some sort of common measurement unit. In reality there are no stable once and for all assessments of the relationship between the analogue and the digital realm. This is being complicated by the introduction of working digital. Is there a way out of this?

In Part 2 of the expert meeting the focus was on feasibility of creating one typology for all sorts of heritage domains and what is needed to realise this.

Creating a generic typology of heritage collections

"Is it possible to create one typology that transcends the boundaries between the different types of heritage institutions?"

To investigate the initial ideas about this, the group was asked to exchange initial thoughts on the question above. On the whole the participants had the idea that it will be possible to devise some sort of standard list of [digital heritage] object types: "Defining the object is a possibility across sectors." Some even take a radical stance: "You should not ask: is a typology feasible? It is a necessity. We need that typology."

In discussing the feasibility of asking staff in institutions to work with the typology a pragmatic consideration was expressed: it cannot be too refined. One attendee said: "If it could be on a high level it might work: "pieces of art" is ok… prints, engravings etc. is much too detailed." We should not ignore the fact that such a generic typology does not exist yet, even though professional collection management has been around for a long time. There must be a reason for that...

Another problem was recognized in the differing realities in the analogue and digital realm. It is clear that determining object types is not a simple issue in the analogue domain. As one remarked: "Some complex manuscripts are in effect archaeological sites… They are compound objects with different versions, interpretations, improvements, and commentaries. It is difficult to count under these circumstances. There are lots of layers." But in the digital realm interrelationships between (parts) of documents become even more complicated. Digital objects are compiled of different types of data. "In the audiovisual domain: it is all-in wrappers." In the library world one speaks of "complex objects".

A related complication is in the imbalance between the analogue objects and their digitized manifestations/reproductions. One analogue heritage object - the example was Rembrandt's painting of the Night watch in the Rijksmuseum - can be digitized in many different manifestations: high resolution images, details, spectrographic images, etc. Even tabular data, covering specific technical measurements may be seen as a digitised version of the artwork. If all of these are taken together as a conceptual **digital unit**, then a generic typology might be possible. So the real challenge seems to be the assessment of such digital units.

One might think that this does not count for simpler objects, like analogue and digital photographs, but this does not seem true. The digital heritage object may of course be restricted to immediate copies, more or less reproducing the features of the original. This way, if you have one painting, you can create one digital image of it. But it is more complicated than that: "even a photograph may have a negative and several differently made prints of it (dark/light/retouched etc.)."

The participants agreed that trying to establish an unambiguous link between the analogue world of cultural heritage objects and the world of digitally reproduced heritage objects is too hazardous. For that reason it was proposed to leave the link between the analogue and the digital realm for what it is: "I think it might be necessary [to have some typology]. And it might be feasible, [but] only if we focus more on the digital experience than on the analogue experience." The digital unit is more or less the same across sectors.

An additional advantage to this approach, focussing on the digital, is that it will not be necessary to create a separate typology for born digital materials.

To conclude: all participants underlined that it is possible to create one typology transcending the boundaries between domains if we start from the digital end, not the physical one. Big advantage is that we also have the born digital materials on board.

The next question posed was more pragmatic in nature:

"Can a generic typology be more than a bureaucratic construct that can also serve institutions, e.g. with collection management?"

As could be expected participants in the meeting thought that any typology should be relevant to the people working in the heritage institutions. One might even go as far as to set the ambition to create a "a definite qualifier list in the Dublin Core descriptor set". Developing the typology should be a bottom-up approach, rooted in and with relevancy for the daily work in the institutions. But it was noted that there is a danger in this as well. From DEN comes the experience from research into born digital heritage (born digital art) that institutions have a tendency to avoid collecting "problematic" object types. It will also be necessary to look at the developments in information management outside the cultural heritage institutions.

Another complication was identified, while returning to the Night watch/Rijksmuseum example. If the full package of digital materials that in some way reproduce or document a masterpiece is considered as a digital unit, is may prove to be impossible to determine the boundaries of what a digital unit is.

Furthermore, it was agreed that museums are organized very differently: some focus on local history, others on art, others on archival materials. When different museums count objects, it happens in different ways. So it will not help to try to harmonize counting procedures by ignoring the different characteristics of institutions. The focus should be on (the value of) the objects. The unit to define is the individual object type.

"What would be the uses of a generic typology? Who would benefit (most) from it?"

In the discussion on specific uses of a generic typology those present noted the difference between benefits for the individual institutions and benefits in a national or European context. The group agreed that it is probably easier to define the usefulness on a collaborative level than on the level of the individual institutions. One participant said: "From the perspective of Europeana knowing how much there is in collections, is very useful information. […] Europeana sends monthly reports to the EU commission about how much there is in the Europeana database. But it is also important to have data about: what is around there, what is in the various institutions."

And what about the value for policy makers? If a generic typology be deployed it may be useful for strategy development on a national scale. It is an instrument that can be used to determine where there are flaws in national heritage programs. Born digital materials are mentioned as a striking case. If a typology is used to measure the progress being made in collecting born digital materials - or the lack of it! - that would certainly help policy makers in state bodies developing national action plans.

To resume, the following uses of a generic typology were identified:

a. questionnaires, such as the ENUMERATE one

b. collections management per institution

c. collection and selection policy on a national scale

d. improving search strategies (e.g. in Europeana)

e. (reporting to) policy makers, collaboration activities, funding

"Which level of detail can realistically be expected from a typology that can be used by all heritage institutions?"

The group discussed a born digital typology produced by DEN during a Dutch survey on born digital heritage (in 2009). One of the outcomes of this particular research project was that lists of object types that are too detailed are disadvantageous for the response rate. More in general the scope of this is confirmed by one of the experts: "If there is too much detail: people in the (…) museum will sigh. (…) Every museum in the Netherlands has its own classification of materials and any museum would have difficulty counting objects using a general typology."

Interestingly the DEN list of born digital object types elicited quite a few responses from the group of experts. Specific categories of born digital materials - e-mail files, AutoCad files, etc. - were missed, since these are in the DEN list classified under high level headings (personal archives, digital 3D designs or reconstructions of objects and buildings, etc.). A suggestion to overcome the problems identified is to develop a classification instead of a typology. The hierarchy of a classification can be used to control the different dimensions and relationships that exist between digital units. It was agreed during the meeting that "Classification" is the better term to use from now in the ENUMERATE project.

After a short break an inventory was made of other research initiatives that can be used as a foundation for the proposed classification of object types. Those present are invited to continue sharing additional references with the ENUMERATE Team. References will be published on the ENUMERATE Delicious list at: http://delicious.com/enumeratesources

How to deal with heritage collections that cannot be counted easily

In the NUMERIC project and during the formative process of the first ENUMERATE Core Survey four types of heritage provided difficulties: archival records, newspapers, monuments and sites, and born digital heritage. The problems with the former three are caused to a significant extent by the lack of harmonisation and standardisation in collection management approaches. This makes it very difficult to ask questions regarding the size and growth of the physical collections. The solution for this problem lies in the approach discussed earlier in the meeting to count from the digital end, not the physical one. What is needed, is a generic concordance from the digital to the physical (x numbers of scans relate to x meters of shelves or x issues of newspapers). Such generic concordances do not exist yet, and ENUMERATE needs to look into the options to develop them in the project.

About born digital heritage, the additional question that emerged was: When is something considered to be heritage? Aside from digitising their heritage collections or collecting born digital heritage, many institutions also posses digital research data and other forms of documentation that provide context for heritage collections. This data will be lost when not properly preserved.

The attendees of the meeting had no reservations about the importance of these contextual data, however, two questions arose: Is the contextual documentation surrounding an object of cultural heritage part of the digital (or digitised) object? Is the contextual documentation cultural heritage in itself and, if so, should it be preserved by the institution? Understandably no definitive answer could be given to these questions, but at least for research data a recommendation was made. The attendees recommended that research data should be part of future renditions of the ENUMERATE survey. In this manner institutions can decide for themselves whether they consider this to be cultural heritage.

Other suggestions for future ENUMERATE surveys

In the tour de table the participants gave suggestions to the ENUMERATE Team. These varied from a warning for oversimplification of questions on cost data, to remarks about survey fatigue and the need to clearly explain what is the use of doing the survey: "It should be clear what's in it for us", to emphasising the need to raise awareness of the long term sustainability in heritage organisations.

Some final points:

We are only witnessing the first generation of digital output. Understanding the planning of long term preservation is vital to move forward with digital collections.

To approach the topic of measuring size and growth of digital heritage collections from the digital perspective is the way to go forward.

You will hardly get any response if you just send questionnaires around. Go to the museums when doing the survey and assist them in person with responding! That way you get very consistent output.

Through future questionnaires it would be valuable to learn: How do institutions organize their workflow? How do they implement policies? How does management take it up? At present digital asset management is not considered to be a core business!

More attention should be paid to the outreach side of digital heritage collections. How much benefit do you get? How many people know about your collection? Is your government happy?

Investigate the barriers caused by legal issues: how much work does it take to settle these?

Keep questions about the built environment separated from those dealing with heritage objects in museums, libraries and archives. The new EU project EU-CHIC would be a good partner to liaise between the two worlds.

Take storage problems (may change), public private partnerships in relation to the ownership of images; and the different demands of different target groups into consideration.

Recommendations

Definitions

1. Promote a common understanding of what cultural heritage is.

2. Define criteria or guidelines of what information objects can be considered as digital cultural heritage.

Context

3. For heritage wide harmonisation, only generic, high level solutions can be applied. The practice across sectors is too diverse to change.

4. Measuring the actual status of digitisation should be related to understanding the total costs of ownership (i.e. including the costs of digital preservation).

5. Understanding the developments in information management outside the cultural heritage institutions is needed in order to create new selection and acquisition policies for digital heritage

Survey methodology

6. The starting point for improving the methodology of measuring digital heritage collections should be the digital realm, not the physical one.

7. Measuring digital collections requires a better determination of digital units (which may be simple or complex).

8. Determining digital units can best be done on the basis of a hierarchical classification that expresses dimensions and relationships.

The cultural heritage field is invited to share references to projects, publications and other initiatives in measuring digital heritage collections with the ENUMERATE Team (e-mail: answers@enumerate.eu). References will be on the ENUMERATE Delicious list at: http://delicious.com/enumeratesources/

8.4 *Measuring the Cost of Digital Cultural Heritage* (Report on Specialist Meeting 2, London, March 30, 2012)

Date: **30th March 2012**

Venue: **CILIP offices (Charter East room), 7 Ridgmount Street, London, UK**

Invited experts:

1. CENL/The European Library: Hans Petschar
2. Czech National Library: Adolf Knoll
3. ICA: Alison Macdonald
4. National Maritime Museum: Richard Ferguson
5. Panteia: Natasha Stroeker
6. Panteia: René Vogels
7. PrestoCentre: Richard Wright
8. MH-Strategy: Max Hammond
9. University of Amsterdam: Trilce Navarrete

Participating on behalf of ENUMERATE:

10. Collections Trust (United Kingdom): Gordon McKenna
11. Collections Trust (United Kingdom): Nick Poole
12. Collections Trust (United Kingdom): Katie Smith
13. DEN (Netherlands): Gerhard Jan Nauta
14. DEN (Netherlands): Marco de Niet
15. Digibis (Spain): Jesús Domínguez
16. Digibis (Spain): Maribel Campillejo
17. Stiftung Preußischer Kulturbesitz (Germany): Monika Hagedorn-Saupe

1. Summary

This 2nd ENUMERATE specialist meeting was dedicated to improving the methodology to measure costs of digitisation. The goal of this meeting was to discuss the need for such type of research, the usefulness of a common methodology and the difficulties encountered so far in ENUMERATE and the previous NUMERIC project.

The meeting started with a presentation of the ENUMERATE Thematic Network and its ambitions, followed by a concise overview of related previous and current initiatives on monitoring the cost of digitisation.

On the whole participants agreed that it would be artificial to separate up-front costs from on-going costs and it would be equally artificial to discard born-digital materials and all sorts of acquisition of digital materials where the digital file is supplied by others than the individual institutions.

Furthermore, in any statistically sound follow-up research - in particular in the ENUMERATE Thematic Survey, planned for the second half of 2012 - an annotated overview of clear definitions of costs should be available.

Caution is needed when extrapolating costs queried in the individual institutions to the national level. It will be difficult to prove that the results are statistically valid, because so many variables may have an influence on the figures. But this does not mean that the extrapolation of costs is not worthwhile to pursue. Here too clear definitions and a convincing segmentation model are a prerequisite for further research. What are the characteristics that are decisive for the costs? It is clear that not all cost factors can be easily extrapolated.

"So the model we are getting into here is a sort of catalytic series of layers. We are talking about object variables, we are talking about institutional context variables, and we are talking about the national context variables." The challenge will be to define what these variables are and to quantify them. There will be differences in between digitisation projects. And such differences may be quite surprising, as in some cases where you (as an institution) are not allowed to involve a third party contractor.

The suitability for the ENUMERATE framework of four recently developed models was also discussed. These models are: Total Cost of Ownership; the Digital Content Life Cycle; Workflow Analysis; and the Supply-Chain Model for Digital Cultural Content. Taking all into consideration of these four models the majority of the group favoured Workflow Analysis most. But serious reflection should be given to defining how broad or how narrow the concept of 'workflow' is taken. It would be a mistake to stick to the traditional digitisation workflow, only mapping out the conversion process. Born digital content may enter a heritage institution at points that were previously unimaginable.

The final topic discussed was the usefulness of cost calculation tools. A general concern with existing tools was that they usually assume an institution starts from scratch with digitisation activities. Furthermore existing cost calculators tend to cover only a subset of the aspects of creating and sustaining digital collections, whereas heritage professionals are getting more and more aware of the cost variables, so the models underlying cost calculation tools tend to become more and more complicated. Generalisations and normalisations are needed to compensate for this. A balance between the specific and the generic can be found, when a calculator is used to collect costs, not predict costs in a generic way. Case based working, it was agreed, is the only realistic way forward.

At the end of the meeting the specialists were asked for suggestions for future ENUMERATE surveys.

2. Background

The primary objective of the EU-funded project ENUMERATE (2011-2013) is to create a reliable baseline of statistical data about digitization, digital preservation and online access to cultural heritage in Europe.

The ENUMERATE Project will conduct three surveys over the course of 2012 and 2013. These surveys are aimed at the following main topics:

- Growth of Digital Collections (supply);
- Usage of Digital material (demand);
- Costs of Digitisation (economics);
- Digital preservation practices (sustainability).

ENUMERATE builds on the results of the NUMERIC project (2007-2009), a ground-breaking initiative to create a framework for the gathering of statistical data on digital cultural heritage. ENUMERATE will improve and refine the methodology from NUMERIC and will bring the data online for re-use.

In order to evaluate and improve the methodology of the ENUMERATE project four specialist meetings are scheduled to take place in the first half of 2012. In each of these meetings one of the main topics of the ENUMERATE survey will be discussed. The outcomes of these meetings contribute to the improvement of the statistical framework of ENUMERATE.

3. Setup of the meeting

The 2nd specialist meeting was dedicated to improving the methodology to measure costs of digitisation. The goal of this meeting was to discuss the need for such type of research, the usefulness of a common methodology and the difficulties encountered so far in ENUMERATE and NUMERIC. The Specialist Meeting also took on board the recommendations from the SIG-STATS (EU Special Interest Group on Cultural Heritage Digitisation Statistics).

Regarding the measuring of costs of digitisation the ENUMERATE survey is intended to address questions like:

- How much did the digitisation of cultural heritage in Europe cost in the past year (2010)?
- How much of the regular institutional budgets are used for digitisation purposes?
- What is the contribution of the EU and national governments to digitisation of cultural heritage?

- What is the investment of private parties in the digitisation of cultural heritage?

- What is the contribution in time of regular staff or volunteers to digitisation activities?

- How much investment is needed for all planned digitisation activities in the next 5 years?

Specialist Meeting 2 was meant to help improve the methodology needed for answering these and other questions.

To structure the discussion, the meeting addressed three main issues:

a. Usefulness of measuring costs of digitisation of cultural heritage

What are the main motivations for memory institutions to record the costs of their digitisation activities? And why would it be useful to compare them among themselves and across the various heritage domains?

b. Feasibility of a common methodology to calculate costs

Calculating costs for digitisation projects is a very complicated issue because of the many variables, yet for the management of institutions, policy makers and funding agencies a better insight in these costs would be profitable. What are the chances of reaching some level of standardisation across the heritage domains to calculate the costs of digitisation? The SIG-STATS recommended to make a distinction between a few global specifications of costs that can be gathered and more detailed data, perhaps in a separate survey on costs. Is this feasible, and if so, how?

c. Cost models to predict future costs

The third topic addressed considered the usefulness of existing cost models, especially in relation to predicting future costs of digitisation (time span: 5 years). Is there a cost model that has preference above others? Or is there a need to create a new cost model to do so? What is the best way forward for this?

4. Previous and current related research on the cost of digitisation

The meeting started with a presentation of the ENUMERATE Thematic Network and its ambitions, followed by a concise overview of related previous and current initiatives on monitoring the cost of digitisation.

a) NUMERIC

In preparation of the NUMERIC Study (2008-2009) a thorough *analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials* stands out. Zinaida Manzuch (Vilnius University) reviewed the major monitoring projects in the years 1997-2007. Manzuch's report covers such questions as: How is digitisation funded? What are the digitisation expenditures of heritage institutions? What are the major costs of digitisation? And what are the components of costs and cost factors influencing the amount of expense?

One of Manzuch's findings was that: "None of the reports provided a methodology that was applied for calculating costs. The content of survey instruments sometimes suggests that respondents were faced with the necessity of calculating data themselves."[4]

Manzuch identified five (high level) cost factors:

- Digitisation objectives;
- Collection parameters;
- Quality of the digital surrogate;
- Application of cost reduction mechanisms;
- Technological infrastructure.[5]

Relating further specifications of these cost factors to four digitisation processes (preparation of digitisation, digital conversion and management, post-digitisation activities and project management and operation) enabled her to roughly specify high and low costs. For example the digitisation objective *access* resulted in relatively low digitisation costs as compared to relatively high costs associated with the objective of *preservation*.

---

[4] Zinaida Manzuch, An analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials (2007), p.29.
[5] Zinaida Manzuch, op.cit., p.31.

The focus in these earlier research reports was primarily on a rather narrow concept of digitisation, as is clear from the cost factors identified by Zinaida Manzuch. These are about the conversion process as such. Little or no attention is paid to models that are currently gaining importance in the management of digital heritage collections, such as Total Cost of Ownership (TCO).

Manzuch's study also made it clear that there is a bias in answering depending on who is asking the question and why they are asking the question.

Regarding the NUMERIC Study itself, it proved to be difficult to assess the reliability of some of its outcomes, as a few example tables with aggregated results from the NUMERIC Study Report clearly show.

- A table about the financial resources identified shows very different figures for digitisation budgets as a percentage of the total budget of heritage institutions for different types of institutions. For example, the digitisation budget in Broadcasting institutes was estimated at 0.3% of the institutional budget, whereas for Archives/Record offices this was estimated to be 4.5%.[6]

- In another table the median reported digitisation costs per broad object type (images alone, text & images, text alone, audio, etc.) varied between improbable extremes. For example the digitisation of one hour of film was estimate to be €128.89 in A-V or film institutions and only €16.75 in National libraries. Equally remarkable: in A-V or film institutions the estimated price of digitising Text & images (per page) was much higher than the price of digitising (per page) Text alone, whereas in Archive/records offices it was the inverse.[7]

b) SIG-STATS

The Special Interest Group on Cultural Heritage Statistics (SIG-STATS) noted that substantial variance was found in the Numeric survey concerning the recorded costs of digitisation. Here are some explanations for this:

- The way questions about expenditure were phrased may have resulted in too much ambiguity;
- The inclusion or exclusion by respondents of 'invisible costs', such as staffing costs, costs of making policy or project plans, etc.;
- There have been different persons from the same institution answering the survey questions, or different types of persons across the institutions;
- The complexity of the organisations could vary considerably.

SIG-STATS concluded: It is not justified to make generalisations about digitisation costs solely based on the responses in the Numeric survey. But SIG-STATS also proclaimed: "calculating costs for digitisation projects is (…) complicated (…), yet for the management of institutions, policy makers and funding agencies a better insight in these costs would be profitable."

c) Curtis+Cartwright Report: *Understanding the costs of digitization*

An indication of the prevailing cost factors in some big digitisation projects comes from a research report entitled "Understanding the costs of digitization", by Max Hammond - one of the invited specialists at the meeting - and Claire Davies.[8] The main cost factors identified in this research report are project management, content capture, metadata generation, procurement, IPR and service delivery. The report is an annotated breakdown of these general costs, with broad indications of how such cost items may vary according to the particulars of a digitisation project.

---

[6] *NUMERIC Study Report* (2009), p.59.
[7] *NUMERIC Study Report* (2009), p.59.
[8] Max Hammond and Claire Davies, *Understanding the costs of digitisation: detail report* (2009).

d) Collections Trust. *The Cost of Digitising Europe's Cultural Heritage* **(2010)**

Nick Poole introduced the cost study he has written for the Comité des Sages, whose task was "to make recommendations to the European Commission, European cultural institutions and any stakeholders, on ways and means to make Europe's cultural heritage and creativity available on the Internet and to preserve it for future generations"[9]. The underlying questions were: How can we increase the pace of digitisation? And how much would it cost to digitise all of European cultural heritage? CT's task was strictly to answer the second question, but in fact the mission entailed much more than "simply" calculating the cost of creating the digital surrogate. It's not the scanning costs that count - the creation of the digital counterpart - what counts is having of digital collections.[10]

Collections Trust set out to answer the question by posing three sub-questions: How many museums, libraries and archives are there in Europe? How many objects do they have in their collections? And how much would it cost to digitize all these objects? To make answering practicable all objects were assigned to a few very broad object classes: simple and complex books; simple 2D; complex 2D; simple 3D; complex 3D; simple and complex audio and video.

The outcomes of all calculations were presented in three sets of figures, both with low and high-end estimates:

- The costs of digitisation done in-house and project-based;
- The costs of digitisation when it is outsourced;
- The costs of digitisation in public-private partnships (PPP).

The cost of digitisation in PPP was estimated to be lowest. In-house, project-based digitisation was rated as the most expensive solution. There were estimates based on doing the job at different paces. And although it was not strictly part of the mission, one of the goals of the study was to make cost of ownership visible.

On request of the Comité des Sages the costs of digitising heritage materials were compared with other public investments. In this report the digitisation of a percentage of all available books in EU libraries, or of all historic photographs in EU memory institutions were compared to the cost of delivering 100km of main road in Europe, or the cost of producing a Joint Strike Fighter plane.

An interesting but puzzling figure was what part of the European heritage collections should be considered not appropriate for digitisation. Another interesting topic had to do with "accelerator issues": the longer the digitisation process runs, the faster it gets; your workflow becomes more efficient and the cost will decrease.

Part of the assignment was the design and creation of an online calculator tool. The tool can be accessed from the Collections Link website.[11]

The political message of the study was: digitisation of cultural heritage is big money, but not 'BIG big'.

e) ENUMERATE Core Survey

In the ENUMERATE Core Survey, which ran from January to March 15th 2012, only some broad questions concerning the cost of digitisation were included. The respondents were asked to give an estimation of the total budget spent on digitisation, the total number of staff involved, the main sources of funding and the factors included in the digitisation budget. For this, the following entries could be selected: Staff costs, Taxes, Capital costs, Equipment costs, Selection costs, Digital conversion costs, Metadata enhancement, Metadata creation, Professional fees, Rights clearance, Digital preservation costs. At the time of the meeting, the results of the survey were not yet available.

---

[9] http://ec.europa.eu/information_society/activities/digital_libraries/comite_des_sages/index_en.htm

[10] "The Cost of Digitising Europe's Cultural Heritage: A Report for the Comité des Sages of the European Commission", prepared by Nick Poole, the Collections Trust, November 2010: http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/annexes/digiti_report.pdf

[11] http://www.collectionslink.org.uk/discover/sustaining-digital/723-digitisation-costs-calculator

f) ITHAKA S+R survey: *Sustainability of Digitized Special Collections*

A final example presented was a recent survey - the results of which still had to be published at the time of the meeting - amongst North American higher education libraries having special collections - the ITHAKA S+R survey on the *Sustainability of Digitized Special Collections*. In this survey the issue of organizational granularity is addressed. The survey is intended to "provide a detailed portrait of how libraries and cultural institutions are managing their digitized collections; offer useful financial data on what libraries and cultural institutions are spending not just to create, but to sustain these collections; shed light on the various sources of support - financial and non-financial - for these activities; [and] highlight some projects that provide valuable examples of strategies and tactics of use to others facing similar issues."[12] The ITHAKA survey was broken down into three separate questionnaires, to be used at three different levels in each institution:

- The Institutional Perspective: to be explained by the director of the institution in a short (10 minutes max.) questionnaire.
- Collections in the Aggregate: a more detailed questionnaire on the institutional collections, taken together (60 minutes).
- Specific Collections: investigated the specifics of single collections. The questionnaire was designed for project leads (45 minutes approximately).

5. Comments on the previous research

After the presentation of these research activities, the participants felt the need to arrive at clear definition of digitisation.

NUMERIC used the IMLS definition: "...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."

SIG-STATS recommended to "...use the more generic Wordnet definition for 'digitisation' as a general starting point: "conversion of analogue information into digital information", and from there make explicit distinctions between 'digital descriptions' (or 'metadata') and 'digital reproductions' (or: 'representations') in future surveys."

The group agreed that most problematic in both definitions is the notion of "conversion". It is evident that born digital materials are not covered this way. It would be best make a clear distinction between 'digital' and 'digitisation/digitized' from now on. What ENUMERATE is presently trying to do is to provide a cost model for the transition the institutions are in, but the project also need to look at the costs 'post transition'. In that sense, it might be better to consistently talk about the cost of digital collections instead of digitised collections.

In relation to this, the *up-front costs* and the *on-going costs* as distinguished in the ITHAKA S+R Survey were discussed. Ithaka defines these as follows: Up-front costs are "expenditures for first-time creation: costs incurred for the creation of the digitized collection up to the point of public launch." On-going costs are "expenditures for on-going maintenance, enhancement, and preservation post-launch."[13] Not all participants found the distinction clear or useful. There is also the difference between 'costs' and 'expenses' ('out of pocket').

It was noted that up-front costs do not necessarily equal the actual analogue to digital conversion process ('digitized collection'); still this is on the whole a substantial component of these costs. The Ithaka distinction may be useful in relation to funding agencies: most of them will only support the actual conversion process and not the long term maintenance of the digital collection. However, professionally operating institutions would not separate these two. It was also noted that there are funding agencies that won't provide the up-front costs anymore, since the creation of digital collections are considered to be the regular business of the heritage institutions.

---

[12] http://www.ithaka.org/ithaka-s-r/research/sustainability-of-digitized-special-collections (accessed March 25th 2012).
[13] http://www.ithaka.org/ithaka-s-r/research/sustainability-of-digitized-special-collections

Some other issues related to doing a survey on the cost of digitisation that were addressed in the discussion:

- Expecting only one single response from a larger institution is problematic. Institutions are often segmented and respondents cannot easily cover the large institutions. Some know the costs of staff, equipment etc. Some know all costs. And some cannot specify any costs at all
- If you compare digitisation costs, you will not be able to get sound results over all if one interviewed person is (implicitly) talking about low quality digitisation (e.g. digitisation of text as an image) while another considered the outcomes of digitisation actions aimed at high quality output (e.g. digitisation of text as hand-corrected, machine readable text).
- Be aware that the type of budget employed for digitisation can have an effect on the specification of costs: "Some projects have a specific (recurrent) budget, in others some (one-time) budget - possibly more or less independent of the envisaged digitisation project - is found somewhere in the organisation."
- The digitisation objectives should not be overlooked: In the audio-visual world preservation is the main driver. From a preservation point of view you do other things (differently prized) than when you are merely digitising for access.
- In research on cost issues, there should be a clear distinction between recording actual costs and projecting future costs.

To summarize: on the whole participants agreed that it would be artificial to separate up-front costs from on-going costs and it would be equally artificial to discard born-digital materials and all sorts of acquisition of digital materials where the digital file is supplied by others than the individual institutions.

Furthermore, in any statistically sound follow-up research - in particular in the ENUMERATE Thematic Survey, planned for the second half of 2012 - an annotated overview of clear definitions of costs should be available.

6. Usefulness of measuring the cost of digitisation of cultural heritage

After this short historical overview, the group was asked to comment on the usefulness of measuring the costs of "digital collections". A few questions were formulated to structure the debate.

What are the biggest benefits of having more precise information about costs of digitisation (and why not accept costs of digitisation as part of a wider budget)?

The group identified several entities to which this kind of information would be useful, especially in relation to accountability: : the institution itself, the heritage domain, national governments, the European Commission and (other) funding agencies.

Common for all entities is the benefit of an increased awareness of the importance of reliable cost estimates in a rapidly changing world: the heritage domain comes from a situation where local (physical) collections were maintained and extended for local audiences. The costs associated with that are fairly simple, at least in theory. With the digital, things get different. The creation of digital collections in a networked environment becomes a joint effort, irrespective of where the actual walls of museums, libraries and archives stand. Access to digital collections becomes global and so may be the audience. The costs associated with this change and are far from easy to specify.

If we zoom in on the entities, there are several reasons why better data about costs and investments in digitisation are useful.

Institution

On the whole managers lack evidence to decide on priorities and what the return on investment could be for the institution. "We need toolkits to get best guess metrics." Organisational changes resulting from the digital developments make it necessary to explain a vision: where do you want to be with your institution?

In addition to this, it is troublesome, without further specifications or guidance, to assess the costs of activities in wider digital infrastructures and frameworks, like for instance the costs of delivering data to Europeana. More precise information from institutions with experience in the field would be of value.

In general, there is a lack of understanding of the digital transition. The associated risk is spending money on digitisation without a clear picture of the desired targets. Why are you digitising? You can't do the whole thing, because you cannot afford it.

The more you digitise, the more you will have to take future costs into account. We often do not know the consequences of our digitisation activities. Understanding costs now will help decrease future costs, e.g. the cost of digital preservation.

Finally, the institutions need this kind of data for their accountability to society. Especially in Europe, where museums, libraries and archives are for a large part funded from public money, it is reasonable that they offer insight into how the money is spent.

Heritage domain

For the heritage domain, as a whole, more precise cost information could be a driver for collaboration. There is a clear shift from institutional to national ways of building digital collections and services. Having a harmonised vocabulary on measuring costs and maybe even shared practices is important for strengthening the sharing of resources and exploring opportunities for economies of scale and synergy.

National governments

Attendees noted that at the national level (and international levels for that matter), there is a shallow focus on mere quantities of digitised objects as an indicator of success in national policy. The topic of digitisation is oversimplified. Creating a baseline for evidence-based policy would be an important step forward.

European Commission

On a European level a lack of intelligence about the cost of digitisation is impeding informed decisions about where and how much to invest in (digital) cultural content - in line with the EU 2020 strategy.

This lack of intelligence also obscures statements about return on investment from European research and other programmes. For the European Commission, (as it is for national governments) Europeana is a success if measured in terms of mere quantities, but estimating the cost of building the Europeana database is only part of the problem. The model is/should be changing into a model of value creation. However, it is a well known fact that the Commission doesn't want complexity.

Funding agencies

The final entity that would benefit most from better intelligence about costs of digitisation would be the various funding agencies at all levels (local, regional, national, and international). Benchmarking costs would be a good approach to assess value for money for funding agencies and tenders.

What cost figures are relevant?

A general remark, which surfaced several times during the meeting, is that the discussion about costs of digitisation cannot be separated from discussion about the benefits of digitisation. There is a moment of appraisal when you decide whether or not you want to keep cultural heritage materials. This is clear, for instance, in substitution digitisation, where you throw away originals or let them deteriorate and only keep the digital surrogate. What's being balanced here is the cost of digitisation versus the cost of having to preserve an analogue original. So input about the relevancy of specific cost figures cannot be made without paying attention to the value that is being created with the investments.

This is all the more important with born digital materials: Several attendees stressed the fact that the digital heritage landscape is rapidly changing. The old situation, in which individual institutions take care of the digitisation of their local collections and all the workflows and management processes associated with and resulting from that, is becoming less mainstream. From the Austrian National Library comes the example of a recent contract with a particular press agency, regulating the acquisition of off the shelf digital photographs covering specific topics in Austrian politics. It is important to know the costs associated with these lateral collection entries. As stated earlier, it would not be sufficient to characterise these developments as 'digitisation'. "Getting a hard drive containing digital stuff in your museum. Is that digitisation?" To be able to identify costs factors, a wider perspective than the traditional approach to digitisation is needed. Nice illustrations of this are crowd-sourcing projects. The audience taking pictures that can be uploaded to some sort of publicly available service - think of Wiki Loves Art - is another example of why the old digitisation paradigm does not suffice anymore to understand cost factors.

With crowd sourcing you have opportunity costs as well. It is not for free. For instance, you need to have a tool to capture the information from the crowd. And once you have captured this kind of "wisdom", you will need systems to guarantee long-term sustainability. Money is needed for that too.

The example of the Tate Gallery was given. The Tate needs to invest substantial amounts to enable user-generated content. These kinds of opportunity costs are often missing in cost queries.

When setting up a survey, it is useful to look at these opportunity costs and other costs that are less visible in a traditional digitisation budget, such as:

- Training costs;
- Cost of delivering data to aggregators like Europeana;
- Copyright costs (clearing);
- The cost of use and reuse of digital collections;
- Aggregated costs of digitisation in a narrow sense;
- Long-term costs of preservation.

7. Feasibility of a common methodology to calculate costs

Is there reason to assume that cost structures in museums are different from libraries or archives? What are the overlaps, what the differences?

To be able to understand similarities and differences in cost structures across different types of heritage institutions, clear definitions are decisive, as a museum may use a different term than an archive while meaning the same thing. This can even happen between two institutions from the same heritage domain. Some cost factors may be typical for specific domains, but in general, the conviction is expressed that the processes in museums, libraries and archives are comparable, all the more when object types are the same. Activities like the preparation for digitisation may be slightly different in different types of heritage institutions, but the retrieval and setup process are considered to be similar. Calculating the cost of digitisation is not about domains and sub-domains, it is about types of objects and the contexts in which they function.

One of the attendees pointed at the role of commercial parties in digitisation. Are these specialised firms working only for museums, libraries or archives? They are not, is the consensus.

Finally it was remarked that differentiating along domains is the wrong message. Sector is actually a generalisation in itself. The reality is that there are many individual institutions that don't operate as a sector.

What cost factors queried in the individual institution can be extrapolated (in a sound way) to costs on a national/EU scale?

"While significant sums have been spent on digitisation to date, it has proven difficult to obtain figures on what has been spent on a national basis so far."[14]

Caution is needed when extrapolating costs queried in the individual institutions to the national level. It will be difficult to prove that the results are statistically valid, because there are so many variables. But this does not mean that the extrapolation of costs is not worthwhile to pursue. Here too clear definitions and a convincing segmentation model are a prerequisite for further research. What are the characteristics that are decisive for the costs to be comparable? It is clear that not all cost factors can be easily extrapolated.

One of the reasons why extrapolation is useful is that large differences across institutions and projects become visible. Because of this, prices of doing the digitisation job as charged by commercial companies might begin to level out. In Spain prices of digitisation work have gone down, partly caused by the fact that many companies are active in the field. This is a valid argument to include external parties in future questionnaires. In the current one, they are left out.

Another reason for extrapolation is to support learning and a knowledge flow on a national scale, e.g. from national institutions to smaller institutions. Roles will be defined differently in different types of organisations, and extrapolation may help to harmonize the definitions by creating a full set of characteristics and comparing differences.

This point led to the argument that the degree of digital preparedness is considered as decisive for measuring and extrapolating costs. The level from which you depart is an important variable. There can be so many other variables in the workflow that comparing the total is less useful than comparing the steps. A suggestion for the thematic survey is to distinguish between beginning - intermediate - and advanced levels of digitisation experience, but all in all it will be difficult to establish some sort of base level across different institutions.

---

[14] Bültmann, Hardy, Muir and Victor, *Digitised Content in the UK Research Library and Archives Sector* (2006) p.99.

Comparing at the EU level will probably point out that there is a competitive component in the EU market. A question that countries will be confronted with is: how mature is the supply of their digital heritage services? Is there a national narrative about the state of digitisation of cultural heritage?

On the EU level, we should make a distinction between the needs of policy makers, who are only interested in generic figures, and involvement of the industry (which needs more granular figures).The industry is for instance an important party in achieving interoperability of information systems. Benchmarking interoperability on the EU-level would be very useful, as there is a lot of locked-in software. The greater the interoperability, the cheaper the digital collections may become.

An alternative to benchmarking the final cost - for which the necessary data may be troublesome to collect in many institutions - may be benchmarking time. In a digital media project at the National Maritime Museum (UK), for instance, the most expensive components in the digitisation process are the photo studio and the documentation department. In many institutions (in the UK) there is not enough awareness of the time (in terms of full time equivalents) spent on digitisation. It is suggested to introduce one or two questions drawing attention to the amount of time and effort it takes to do the digitisation. Normalisation of the awareness of costs would be a big plus.

It is concluded that extrapolation and benchmarking on a national scale is worthwhile, if only to get staff a bit more business oriented, as a counterpart to the passion about the collection they manage. It is pretty pathetic that few institutions are able to answer the question how much the scaling up of their digital services is going to cost.

Will there be such differences across the domains re. The quality of digitisation that a common methodology will not be possible?

A key issue where differences may occur is the quality of digitisation. Would different approaches to quality hinder a common methodology? The problem is illustrated by one of the findings from the NUMERIC Study Report:

"(…) the typical (median) costs reported by different institutional types for text only documents ranged between €0.10 and €0.80 per page; and audio materials between €6.42 and €78.84 per hour."[15]

This variety, is it contextual? Or is that just a fact of the data that we happen to have? Such differences are usually attributed to the fact that digitisation quality may vary considerably across digitisation projects, while in survey questions this is usually not taken into account.

Several attendees emphasised again the importance of clear definitions. Questions must be posed in such a way that there is no risk of mixing e.g. the cost of digitising modern documents with the cost of digitising old manuscripts. And because of poor definitions or poor segmentation of the cost factors, the hidden costs - think of capital cost, the cost of equipment, staff time - are often not included by default.

Costs are not comparable if you don't involve the desired outcome of digitisation (images of text versus OCR-ed text for instance). It is suggested that perhaps ambiguities become less troublesome if questions focus on the results of the digitisation process rather than on the input side. This would be in line with the outcomes of the first ENUMERATE Specialist Meeting: measuring/counting the number of analogue objects to be digitised is more troublesome than measuring the number of digital objects after the conversion.

Some variables (cost factors) that were considered to be decisive for the quality and thus the cost of digital collections:

- Experience
- Infrastructure
- Material types (old/new, fragile)
- Intended use (incl. rights management; long term / short term availability; level of quality)

How do you measure these variables? How do you pinpoint an institution on where they are on the scale of these variables? One way would be to distinguish levels of complexity. Here a reference was made to the "Digital Preservation Capability Maturity Model" (DPCMM), as developed by Charles

---

[15] NUMERIC Study Report (2009), p.4

Dollar.[16] Procedures similar to those of the DPCMM could be part of the survey tools of ENUMERATE: a way of specifying or defining an institutions digital maturity.

The discussion then concentrated on differences regarding intended use. The idea that upon the acquisition of an object you may decide that you intend to manage it only for a limited period of time is not alien to an archive, but to a museum it could be quite a scary idea. The decisions taken about intended use may have large consequences for the cost of digital collections. This effect is amplified in the case of born digital collections. With digitisation you have control over the output; with born digital material the control is much less. An extreme example is born digital art, where the intended use may be defined in terms of perpetuity (improbable), in terms of a life span of 2 to 5 years, or in terms of a more or less thorough documentation of the original work (e.g. a video recording). Such different practices do not make it easy to come up with a common methodology.

It is obvious that there are two perspectives to consider: the organisational-centric perspective, and the object-centric perspective. Discussions about what goes into the process; what state the digital object is in each phase of the process; the transformations that happen to the individual object within this overall transition; what comes out in the end and what we do with it: that is the object-centric perspective. It's a different way of seeing the problem. It is necessary to establish that distinction clearly. The advantage of the object-centric approach is that you can get more specific about the costs.

How decisive are national differences between EU member states?

How determining is the national context, in particular the national economic context, on the actual cost of digitisation within the institutions?

All agree that labour cost is a defining variable. But this variable is modified by another variable, which is efficiency (availability of skills, professionalisation). When measuring the cost of scanning an item, it is evident that values get lower with the passing of time. This can partly be attributed to differences in speed of processing in the digitisation workflows. The variables are the same, but may be different for different (EU) countries. You cannot compare the digitisation of a page in the UK with the digitisation of a page in Serbia. National policy might constrain options that are available in other countries, e.g. allowing institutions to generate own income. So the national legal framework is also a factor to consider. But how do you account for that?

Nick Poole summarized: "So the model we are getting into here, is a sort of catalytic series of layers. We are talking about object variables, we are talking about institutional context variables, and we are talking about the national context variables." The challenge will be to define what these variables are and to quantify them. There will be differences in between digitisation projects. And such differences may be quite surprising and hard to overcome, as in some cases where you (as an institution) are not allowed to involve a third party contractor. But overall, benchmarking can help institutions in the learning curve to become more economic, and showing progress achieved in that respect.

What generic deductions can be made about the differences in costs between digitisation by private companies and in house digitisation?

Private companies provide cheaper services, but there is usually a loss in quality and material handling. Heritage institutions are usually inefficient in their digitisation activities, and thus more expensive. For this reason, the National Library of the Czech Republic has decided to have everything outsourced.

It was noted that there is a difference between outsourcing and public-private partnerships (PPP). In the first there is control of content (by the institution), in the second there may be a loss in control of content, as in partnerships there can be agreements about favoured licensing agreements, which may result in a (temporary) lock down of content.

Institutions tend to fool themselves about in-house digitisation. The actual in-house digitisation cost is often not specified to the full. As mentioned earlier, with outsourced digitisation, the sum total of costs becomes much clearer.

---

[16] "The model is used to measure your own organization's progress against a number of key criteria, and then decide where improvement is needed." Cf. http://digitaalduurzaam.blogspot.com/2011/12/playing-digital-lifecycle-game-dish2011.html

There are various approaches to understanding costs. What model(s) should be the starting point of a cross-domain methodology?

Next, the suitability for the ENUMERATE framework of four models that have recently been developed are discussed. These four models are:

- Total Cost of Ownership: a "financial estimate whose purpose is to help consumers and enterprise managers determine direct and indirect costs of a product or system."[17]

- Digital Content Life Cycle: a widely used comprehensive framework which describes the stages of managing digital content projects and programmes, with the aim of "adopting a good practice approach for each to ensure continued success."[18] The stages identified in the model are Creating, Describing, Managing, Discovering, and Using and Reusing,

- Workflow Analysis: activity based costing, where the activities can be boxed.

- Supply-Chain Model for Digital Cultural Content: "work in progress" by Collections Trust, framing various actors, activities, and objects/instruments (e.g. publishers, scanning, and picture libraries) under broad phases in the supply chain like: Strategy (strategic planning/business models), Extraction (acquisition/selection/etc.), Production (scanning/metadata enrichment/etc.), and Management, etc.

The models are in effect different pictures of the same reality.

The group was asked to assess the appropriateness of these approaches.

The TCO and Digital Content Life Cycle models were considered to be the most problematic. Both models do not reckon appropriately with the role of social media, which changes the flow represented. TCO starts from the assumption that the one who digitises materials is also the one who owns the materials. It would be better to re-label the model into "Total Cost of Stewardship".

The Digital Content Life Cycle is rather exclusively focussed on the conversion process. It is unclear where born digital materials enter into the flow.

Workflow Analysis and the Supply-Chain Model are moving away from the objects and collections and are putting the cost issue into the organisational context.

Taking all of these four models into consideration the majority of the group favours Workflow Analysis most. But serious reflection should be given to defining how broad or how narrow the concept of 'workflow' is taken. It would be a mistake to stick to the traditional digitisation workflow, only mapping out the conversion process

It is expected that conventional practice will be overthrown and this is not just limited to the mere creation and management of digital collections. Take for instance the recently developed procedures of pro-actively collecting press photographs as part of the (digital) collection policy of the Austrian Library: in the old situation curators were specialists in certain fields who used to work for two or three years on one collection. In the new system, they will get thousands of photographs every month. - But it is not only a matter of redefining staff roles. There are also issues having to do with the institutional infrastructure. Both will completely change the library. For the Austrian National Library the Supply Chain model is really the model of a digital library at work.

A point was made that proactive collection building activities are often much cheaper than acting afterwards and acquiring or completing collections. The Austrian Library negotiated the right for giving direct access for scientific purposes, and for long-term preservation. "Which is also an advantage for publishers; because they know that in a hundred years they will have the original files if they need them. So this can be also a cost reductive procedure." It marks the transition to a completely digital library and digital collection process.

It is felt that this is really important: there will be less distinction between the point of publication and the point of archiving. If publication and archiving are getting more and more together, that will become very efficient.

---

[17] Cf. http://en.wikipedia.org/wiki/Total_cost_of_ownership
[18] Cf. http://beta.digipedia.org.uk/wiki/Digital_content_life_cycle

There is a need to give politicians a concept of the changes we are dealing with on an institutional level in relation to the bigger infrastructures. The comparison that comes to mind is the transition from a situation with lots of independent railway networks, into an (inter-)national railway infrastructure. A transition like that takes a one-off/once in a generation, huge capital expenditure. It should be articulated in terms of the benefits of what the integrated rail system will look like from the outside, in terms of economic competitiveness, growth, job creation, etc.

But being convincing is going to be difficult. All the evidence we have is related to a fragmented world. Being able to construct a convincing story, choosing the right model for surveying and monitoring, will be crucial.

8. Cost calculation tools to predict future costs

The final topic discussed at the meeting was the usefulness of cost calculation tools. Several tools have been developed in the past years to help calculate the costs of digitisation activities in general and digital preservation in particular. In the ENUMERATE report on tools for validation and harmonisation (September 2011), such tools have been identified and described. The ENUMERATE partners would like to hear from the specialists at the meeting if they thought that it would be useful to promote one or more of these cost calculators that may help the respondents to provide valid answers to the questions on costs. Or would it be better to construct a new, more generic calculator that would fit in the ENUMERATE methodology?

A general concern with the existing tools was that they usually assume that an institution starts from scratch with the digitisation activity. One needs to understand the logic of the entire model first and then map that to the practice in the institutions. But in daily practices one builds on previous work. The calculators traced usually do not presuppose existing infrastructure.

Furthermore, there are so many factors involved in digitisation, that most cost calculators do not cover all aspects of digitisation. For instance, the cost tool that the DEN Foundation developed does not include costs for training staff.

The bottom line is that we need a template, based on the 'experience factor', to reflect the organisational budget and to identify all the factors and effects. The meeting agrees that cost calculation based on work flow analysis would probably work best. Based on that workflow analysis, some representative cases can be made and with those in mind, a logical model can be constructed that can be tested on its predictive capability.

An issue that needs closer attention is capital cost in relation to equipment. Capital cost is "the opportunity cost of the funds employed as the result of an investment decision" (definition from Wordnet). Because of rapid changes in technology, investments in equipment for digitisation and storage of digital output can be a risk. Can a cost calculation tool include a risk analysis element and normalise this? It is agreed that identifying risks is important, but there is a difference between project risks and cost risks. If a cost calculator is meant for calculating costs of project based digitisation, the risk assessment will be restricted to project factors, such as will the project be finished on time and within budget. The more complex the programme is, the more you need to develop it from the beginning, including the acquisition process. This is part of good programme management.

People are getting more and more aware of the variables, so the models underlying cost calculation tools tend to become more and more complicated. Generalisations and normalisations are needed to compensate for this. A balance between the specific and the generic can be found, when a calculator is used to collect costs, not predict costs in a generic way. It is agreed that case based working is the only realistic way forward and that best guesses are acceptable in such an approach to collect costs. "Become empowered by what you know rather than get frustrated about what you don't know". Monitoring and auditing as projects go along would be good mechanisms to check the validity of the outcomes.

Two final points were made.

Most current cost calculation tools are developed for AV or text based materials, and they are less usable for an object approach, where the object is e.g. a museum object, a building or archaeological find. The result is that costs of digitisation of such objects can be misrepresented in surveys if these cost calculators are used.

Finally, a remark was made that not all institutions would be willing to reveal cost factors to the outside world. If we want a cost model to generate bench marking data, this could be a problem for reliability.

9. Suggestions for future ENUMERATE surveys

The final part of the meeting was a summing up of the outcomes of the meeting and what recommendations can be made for the ENUMERATE Thematic Survey:

Recommendation 1:

Review the terminology and definitions on costs. Use clear, simple, short descriptions in all EU-languages to provide a clear understanding of what you are talking about. However, people usually don't read definitions in questionnaires. So the wording of the questions themselves is vital. Quality assurance of the translations is of the utmost importance.

Recommendation 2:

Distinguish between three levels for the collecting of qualitative information (cf. the Ithaka Survey):

- Collection level
- Institution level
- National level

Recommendation 3:

Scalability is an issue. E.g. maturity models need to have an eye for the right scale. Find a balance between involving both large institutions and smaller institutions. Both are important for understanding costs of digitisation. An online tool will be most convenient for all.

Recommendation 4:

It is vital to have personal relationships with staff in the institutions for advocacy of the ENUMERATE project and its relevancy for understanding and improving your own working practice. We need to ensure that when we approach institutions for this detailed information that we are seen as supportive and it is a collaborative process. In relation to that: ask for a person's decision making's responsibility in the organisation. Involve the director and ask for time to spend on the questionnaire.

Recommendation 5:

Think in the long term. ENUMERATE is a framework to document the transition of the sector. We are heading for a data standard for digitisation of cultural heritage but is it realistic to expect that we can achieve that? More desk research and community involvement will be needed...

Recommendation 6:

Organise a 5th meeting, to bring the various topics that ENUMERATE addresses together such as costs and value.

8.5 *Measuring the Use and Impact of Digital Cultural Heritage* (Report on Specialist Meeting 2, Madrid, June 6, 2012)

Date: 6th **June 2012**

Venue: **Historic Library of the Complutense University; Biblioteca Histórica Marqués de Valdecilla, Madrid, Spain**

1. Participants

Invited experts:

1. Culture24: Jane Finnis (JF)
2. Historic Library Marqués de Valdecilla of the Complutense University of Madrid: Almudena Caballos Villar (ACV)
3. ICOMOS: Rand Eppich (RE)
4. Institut Cartogràfic de Catalunya: Rafael Roset (RR)
5. Ministry of Culture of Spain: Alejandro Nuevo (AN)
6. Ministry of Culture of Spain: Julio Cordal (JC)
7. Museum Analytics: Rui Guerra (RG)
8. National Library of Catalunya: Paquita Navarro (PN)
9. National Library of Spain: Elena Sanchez Nogales (ESN)
10. TEL (The European Library): Alastair Dunning (AD)
11. Universidad Carlos III de Madrid: Elisa Giaccardi (EG)
12. University of Amsterdam: Frank Huysmans (FH)

Participating on behalf of ENUMERATE:

1. DEN (Netherlands): Marco de Niet (MDN)
2. DEN (Netherlands): Gerhard Jan Nauta (GJN)
3. DIGIBIS (Spain): Jesús Dominguez (JD)
4. DIGIBIS (Spain): Maribel Campillejo (MC)
5. Collections Trust (United Kingdom): Katie Smith (KS)
6. SPK (Germany): Inés Matres (IM)
7. Panteia (Netherlands): Natasha Stroeker (NS)

Apologies:

1. KDCS Consultancy: Simon Tanner
2. DIGIBIS: Xavier Agenjo
3. DIGIBIS: Francisca Hernández

2. Background to the meeting

This meeting was the third specialist meeting which focused on measuring the use and impact of digital cultural heritage. The meeting was made up of consortium members and external experts. Experts with the following profile were invited:

- Extensive knowledge of a practical or theoretical nature of measuring the use/impact of digitised resources
- Involvement in user studies or benchmarking activities in digital heritage
- Experience with quantitative surveys among heritage institutions

The meeting aimed to discuss:

- Usefulness of measuring use and impact of digital cultural heritage
- Feasibility of a common methodology to monitor access to and use of digital heritage resources
- Methods and tools to monitor the actual use of digital heritage collections

3. Minutes

Welcome and introductions

MDN gave a short welcome and all participants introduced themselves.

The ENUMERATE thematic network

MDN gave a presentation about the overview of the ENUMERATE thematic network. In this he outlined the motivation behind the project and the aims of ENUMERATE according to the Description of Work.

Introduction to work done so far on measuring use and impact of digital heritage

GJN discussed the groundwork for the NUMERIC project, which ENUMERATE has continued on from. GJN discussed Zinaida Manzuch's 2007 paper, '*An Analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials'*, and the different approaches institutions can adopt to look at how they understand the use and impact of digital cultural heritage.

Discussion on the usefulness of this type of research

After the presentation by GJN, various topics were discussed about the usefulness of this type of research:

- MDN asked the meeting how they felt this research was relevant.

- JF stated that they have previously done work on this through their 'Let's Get Real' project which looked at this issue through adopting a pragmatic process and benchmarking. The starting point for Let's Get Real was new digital modes of communication and how institutions do/can measure that. JF highlighted that it is crucial that awareness of an organisations online presence is embedded and integrated throughout an institution.

- RG stated that there are two current trends of collecting data, the multiple organisational approach or collecting data with a focus on individual institutions. RG also expressed that it is crucial that cultural organisations change the way they are working, ensuring they are transparent about the use and impact of their digital heritage, and ENUMERATE should adopt this approach to present data in a very clear way online. It is important to collect the data, analyse it, retrieve information from it and then, finally get to the point of new knowledge.

- Analytics can be a useful tool for internal leverage within an organisation e.g. to persuade investment within departments, comparing how your organisation competes with its peers etc.

- Benchmarking tools are useful, but the group had different opinions on if institutions should be able to see the results of all the other institutions, or if they should be able to see their own results compared against the results of the total or selected reference group.

- RG stated that ENUMERATE should move towards persuading institutions to present their information/data online in a transparent way, once it is online this then becomes a motivator, also for other organisations.

- Discussion found that there was no consistency in implementations of e.g. Google Analytics.

- AD stated that The European Library is currently opening up a lot of data and they are not always in charge of what happens to it, and it is important to try and understand what the impact of your 'stuff' is having outside of own websites etc.

- ESN stated Nat. Library of Spain (NLS) at first wanted to put everything online. They realise now that it is important to know how people are interacting with their content and the NLS are searching for indicators of user preferences. The NLS information professionals are now looking at questions such as how do our users get at our databases? How do users get to our content? Are they finding what they are searching for? How do they search?

- JF makes the point that not enough time is spent on search engine optimisation (SEO), 60% of all requests for data comes through these channels. The most important factor is the understanding of how your content moves people. JF stated that there is a change in that the interface to an organisations website is actually the aggregated Google search results, not the navigational guidance offered at the website itself.

- RE stated that a lot of monument and site professionals are not willing to spend a lot of time on searching for information and highlights that it is important to pay attention to needs which are below the institutional level.

- RR demonstrated that it is important to understand that different statistics have different usefulness/meanings for people, for example, use of content versus number of visitors. There is a change within his institution, the Cartographic Institute of Catalonia, away from finding maps, to enabling the user to experience the collection; this is reflected in the way the catalogue is (not) being used.

- JD stated that in Spain over the past two years there has been an increase in requests from people wanting to monitor the use of their content and it is important that this ENUMERATE research demonstrates the usefulness of this analysis.

- ENUMERATE could look at the relevancy of use and impact statistics at three levels: the institution, the national heritage domain, and the supra-national (EU) domain.

Summary:

- The key drivers for doing this type of research are:
  - Understanding use;
  - Understanding the value of your collections;
  - Accountability;
  - Raising the relevancy of (new) digitisation work;
  - Making more value from investments already made;
  - Direct future research.
- It is important to have information at various levels, as the need or use of information varies per person or institution.

- Detailed reports should ideally be made for each person in an organisation and should relate to their role, and what these statistics means to them, this stimulates involvement in the data and making better use of it.

- It is important to bring together both quantitative and qualitative research for this project as ENUMERATE is serving individual institutions and the European Union as a whole.

- Much of the discussion focussed on the use of (Google) Analytics, however there are many other possibilities, such as using log files, for instance where API's are involved, etc. As AD noted: "It's not about the use of your platform, it's about the use of your content."

Discussing feasibility of a common methodology

MDN introduced two main topics, cross-domain methodology and methods and tools.

Question: Is there reason to assume that the use of digital collections from museums is different from libraries, archives or other heritage institutions? What are the overlaps, what are the differences?

- FH expressed that libraries use statistics in various ways. Research libraries are likely to measure statistics on downloads/views of material whereas public libraries are less likely to. He stated that it is important to know what is being done with the content, for example researchers use content within their daily work whereas the general public use it as more of a leisure activity.

- Google Analytics can be used as a common methodology in certain specific ways as there are some universal indicators, however, the tool does not cover everything related to the use of collections.

- Some time was spent on discussing user profiles. FH did research in the Netherlands on experiencing heritage collections. Different types of users were considered: All-rounders; Art lovers; Association members; Collectors; Browsers; Cultural family outings/Day trippers; Readers; Not active/Not interested.

- But can we discern specific types of users across different domains in the heritage sector? More nuance is needed here. For example: is it a user who only wants to know the entrance fee? (then: spending a short time on the website is good) ...or does he want to browse etc. (then: a short time spent is bad). Etc.

- Discussion demonstrated that it is not necessarily about discovering a common cross domain methodology to analyse statistics, but about understanding different types of use of an institutions cultural content. MDN agrees with this view, putting forward the idea that we should not classify types of *users*, but rather types of *use*.

- The role of funders needs to be considered as well: different funders will push the institution to focus on different types of use.

Question: **Is there reason to assume that measuring of use and impact of digital cultural heritage is so specific that a dedicated methodology is needed? Or can we implement methodologies from other domains?**

- Cultural institutions are deemed 'trustworthy', valued by their transparency and legitimacy of information. It seems plausible that cultural institutions should relate this somehow to the types of use of their content.

- Cultural Institutions should look at key performance indicators from other institutions.

- Google Analytics could be used to collect specific parameters across different institutions. The idea would be to develop a minimal configuration of Google Analytics and incorporate that within the Thematic Survey.

- Google Analytics is most often used without fine tuned installation and thus raises the issue of its ability to measure impact and value if not properly configured.

- Google Analytics measures the usage of the institutional website and does not cover the use of the digital collections of an institution on other platforms. This raises the issue of if this is advisable when the number of sites where an institution's digital collection can be offered to the public is rapidly growing?

- AD pointed out the restricted scope of Google Analytics and that there should be an emphasis towards creating "a family of methodologies" such as in the TIDSR: http://microsites.oii.ox.ac.uk/tidsr/welcome

Question: Will there be so many differences between the target groups of digital services that a common methodology will not be possible to identify within a single survey? (e.g. research, general public, education)

- NS stated that the project needs to identify what it is looking for with regards to the added value of ENUMERATE.

- RE stated that ENUMERATE will be able to generate more specialised results, Google Analytics can be very broad or very specific but you must know what you are looking for. He also expressed the view that it could be dangerous to rely on commercial organisations to provide a free service.

- AD emphasised his point that Google Analytics should be used only as a baseline for institutions to start analysing their data.

- Opinion was expressed that ENUMERATE could create very basic rules that demonstrate segmentation techniques within Google Analytics to the community. This should include when to use it and what to do to get the results needed. An example of this would be to include 5 common grounds, best practices, how to implement and use it, and what kinds of reports are needed.

- JF pointed out that ENUMERATE figures could also be potentially dangerous to institutions if it highlights that their content is not being used.

Discussing methods and tools

MDN introduced the topic of methods and tools for measuring use and impact.

Question: Is there a preference to separate the measuring of use and impact of digital heritage from the measuring of the use and impact of 'physical' services?

- RG expressed his opinion that in the future the approach to the physical and digital will become more alike.

- AD demonstrated that the focus of ENUMERATE should be on measuring the use and impact of digital heritage, but that the project shouldn't forget where this meets the physical.

- Other options for measuring use and impact mentioned were Piwik, Log Files or TIDSR JISC Toolkit. National Library of Spain experimented with Piwik but the outcome of this was not very successful.

- There is a hesitation in certain countries to use Google Analytics as it gets access to the data being processed.

- JF also pointed out that there are tools for social media analysis, these include Facebook Insights, and MailChimp for analysing email statistics. Facebook Insights could be useful when trying to analyse user generated content.

- Discussion highlighted that it would be useful to make a distinction between interactions with the institution and interactions with the collections.

- Statistics from aggregators are left out of ENUMERATE because the focus is on organisations having memory collections and to ensure that content is not counted twice. However there is an argument for counting the use of content on aggregators that is not hosted elsewhere. As the channels through which content can be published get more and more diverse it may become impracticable to retain clear distinctions.

- The Balanced Scorecard is performance management framework. Using the Balanced Scorecard approach would enable institutions to see a framework of ideas that could be implemented within an organisation. This would also force management to think about digitisation activities.

Suggestions for the next ENUMERATE survey

A number of questions that are relevant for the next survey were presented to the group. These were:

1. What are the high-level priorities for ENUMERATE to get a better understanding of use and impact across the European heritage sector?

2. Should the survey be broken down along specific subdivisions (e.g. target groups, types of digital heritage collections)

3. Who would be the appropriate member(s) of staff to approach?

4. Looking at the current ENUMERATE questions, what questions can be maintained and what questions should be altered?

   - AD suggested that in order to gain a good understanding of the survey results, a set of the same core questions should be asked in 12 months time to be able to analyse any changes that have occurred.

   - AD also claimed that people should be enthusiastic about this as they like statistics, which can be very useful within institutions.

   - FH demonstrated that it is important for the project to focus on how to boost participation in the survey. A way of doing this is through national coordinators and social involvement.

   - The survey should include a question similar to 'Do you use Google Analytics or something else? If not, then why not?'

   - The survey should be directed towards the policy officer at an institution.

   - It is also important to consider what success looks like when creating the survey, will the questions you are asking give the results that will enable this success?

- It was highlighted that it may be difficult to ask institutions for usage stats at a certain point in time. Instead a better approach would be to educate institutions in collecting the relevant data and to do this over a large time period. This approach has been adopted through Culture24's "Let's Get Real" project which organised sessions with a selected group of institutions.

- JF wondered whether it would be easier to consult institutional websites and find out what is accessible there as opposed to asking an individual from and institution to answer questions. This would remove the issue of identifying the correct person to complete the survey, encouraging people to complete it and any issues surrounding differences in interpretation, first figuring out what can be done automatically.

- The survey should also look at questions relating to mobile access and social media use (and how this is measured).

- The survey should add in a subject which relates to the classification of types of use (not users).

- The survey should look at what institutions should do after digitisation and how they can bring their information to the public in the most effective way.

- If the survey wants to find out information from monuments or heritage sites it may be difficult as their data is generally aimed for a specific audience and there tends to be little reuse of the specific information collected within the databases.

- GJN stated that ENUMERATE should find a way to combine the approaches of Let's Get Real/Museum Analytics (i.e. a prolonged working together with memory institutions during a longer time span) and the more conventional framework where a questionnaire is used - at one point in time - to query memory institutions (shorter time span).

Any other business

- Participants to email DEN with any further ideas for survey questions. Expenses form sent to KS.

8.6 *Monitoring and Measuring Digital Preservation Practices* **(Brainstorm session, The Hague October 5, 2012)**

Date: **5**[th] **October 2012**

Venue: **Digitaal Erfgoed Nederland (DEN), The Hague, The Netherlands**

Participants:

1. Panteia: Natasha Stroeker
2. PACKED, Flanders: Bert Lemmens
3. National & University Library, Slovenia: Alenka Kavčič-Čolić
4. KB, The Netherlands: Marcel Ras
5. KB, The Netherlands: Jeffrey van der Hoeven
6. DEN: Marco de Niet
7. DEN: Jeanine Tieleman
8. DEN: Gerhard Jan Nauta

1. Introduction

Purpose of the brainstorm session on Digital Preservation was supporting methodology development for the ENUMERATE survey framework. In the description of work for ENUMERATE the topic of monitoring digital preservation is introduced in the following paragraph:

The key factor impeding the widespread adoption and implementation of Digital Preservation Policies is no longer technical knowledge. Over recent years, a considerable corpus of research has generated detailed specifications and guidance to promote sustainable digital assets. Instead, it seems that the key factor is intelligence – there is no overall picture across the EU, or within Nations, about how Digital Preservation policies are being adopted and implemented. There is no statistically-valid basis upon which to benchmark Digital Preservation, or through which to normalise or harmonise issues such as cost and format.

ENUMERATE will extend the NUMERIC model in order to address this intelligence gap by generating the beginnings of a longitudinal dataset about the following:

- *The adoption of Digital Preservation policies at an institutional level*
- *The growth rate of digital content held by cultural institutions*
- *An assessment of the stability of that content and the quality of its preservation*

As a preparation to the meeting all participants were asked to reflect on the type of intelligence that would be needed, both in order to support individual institutions in the management of their collections - an essential incentive to participate in any kind of survey effort - and also, in the long run, to support national and EU governments in adopting and implementing digital preservation policies.

Some guidance questions were distributed; to be discussed at the meeting (Is it possible to answer these questions? Are these questions suitable in a focus group?):

- What are the main motivations for memory institutions and other stakeholders (e.g. policy makers, funding agencies) to monitor the status of digital preservation?

- Are the needs for monitoring preservation practice in the various heritage domains (archives, libraries, museums) comparable?

- How useful is it to compare national strategies for digital preservation?

- Is the amount of time and money needed to monitor digital preservation in individual institutions in proportion with the possible benefits to be gained?

- What kind of monitoring data are feasible? To what extent can these data be quantitative in nature?

- Is there a reason to assume that the cultural heritage sector requires a specialised, domain specific methodology, or can/should a wider, more generic methodology be applied?

- What can we learn from neighbouring domains (e.g. science, government)?

- What is the usefulness of the existing cost models on digital preservation for cultural heritage institutions?

- Can one specific cost model be regarded as the best to fit the needs of the ENUMERATE project?

- What other tools, methods and methodologies regarding preservation issues can be recommended for inclusion in the ENUMERATE statistical framework?

- Can the implementation of specific standards for digital preservation (OAIS, PREMIS, DRAMBORA, and TRAC etc.) be measured? If so, how?

A second request to all was to review a few questions on the topic of digital preservation that were part of the ENUMERATE Core Survey (February-March 2012). These questions were distributed together with an analysis of the answers given by the institutions that participated in the survey.

Finally all were asked to think about existing initiatives in this field in order to share them with the group.

2. Outcomes

Outcomes from the 5th October meeting relate to the Core Survey and Thematic Survey.

Core Survey on digital preservation

This question in the Core Survey in digital preservation should be reconsidered:

- Is your organization included in a national or institutional digital preservation strategy?

Actually there are two questions:

- Is there a national strategy on digital preservation in your country?

- Is your organization included in that national strategy on digital preservation

The same holds for the question on the digital preservation infrastructure.

Question:

- What type of staff is involved in the digital preservation process in your organisation (IT, Collection etc.)?

Thematic Survey on digital preservation

Suggestions to OAIS:

- Are you familiar with OAIS, as a model or strategy of digital preservation?

- Do you understand how it works, what is needed etc?

- Have you implemented OAIS in your organization?

- How have you implemented OAIS in your organization?

** How many repositories are there in Europe in Cultural heritage sector that meet the mandatory requirements of the OAIS ISO standard, ISO standard 14721 (2003)

** How many e-deposits have done an audit based on the TRAC ISO standard?

(This question may come too early?)

Standards for digital preservation (to be decided: is this to be tested in the Thematic Survey or should it be included directly in Core Survey?):

- Do you know / are you familiar with PREMIS?

- Do you have your data available in PREMIS?

- In which way does this help your organisation? [think of a better formulation]

- Ask for all four standards: OAIS (see above), PREMIS, DRAMBORA, TRAC

** Do you use a system from another institution to preserve your digital collections?

Ad National coordinators

At the start of the thematic survey we will ask the National Coordinators if they want to cooperate with the Thematic Survey and which parts they prefer to cooperate in. In this phase we can also ask the national coordinators about the state of the national strategy on digital preservation in their country (**Which countries have an established national preservation strategy?)

- Is there an established national preservation strategy in your country?
- What is the status of this strategy? and:
- Is there an established national digital preservation infrastructure in your country?
- What is the status of this infrastructure? Etc.

8.7 *Questionnaire on the SIZE of digital heritage collections*

Summary

This questionnaire was developed by Natasha Stroeker (Panteia) and Gerhard Jan Nauta (DEN). At the core of the questionnaire is a list of object types that may be represented in digital heritage collections. Since the aim was to assess the size of both digitised analogue and born digital collections (combined) and since such a list could not be traced in earlier research and surveys, the list had to be developed afresh. The list used here is a combination of the NUMERIC/ENUMERATE Core Survey list of object types and a list of born digital object types that was developed in research on born digital heritage in The Netherlands, initiated by DEN in 2009. Accompanying questions are selected to put the results of assessing collection size and growth in context.

Quantifying digital collections is done in three slightly different ways. The initial plan was to ask each of the selected institutions to test each of the separate approaches to measuring collections size. To minimise the risk of impairing the willingness of respondents it was decided reduce the number of approaches to be tested to two per heritage institution. Purpose of this approach is to assess the variance that results from using different methods of counting. The institutions in this part of the TS will be asked to evaluate the practicability of the different methods, which may perhaps lead to varying conclusions about the appropriateness of different methods in different subdomains of the heritage domain.

The National Coordinators in 7 different EU member states have selected the institutions (10 to 15 institutions per member state).

SIZE

Document history

Date	Author	Status
5-2-2013	Natasha Stroeker (Panteia), Gerhard Jan Nauta (DEN)	Draft #1

ENUMERATE SURVEY ON THE <u>SIZE</u> OF DIGITAL COLLECTIONS

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

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

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

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

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

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

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

Best Regards, the ENUMERATE Team

Natasha Stroeker: enumerate@panteia.nl

=== questionnaire ===

SECTION 1/5 - ORGANISATIONAL INFORMATION

[1] Name of institution / organisation *:

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

[2] Type / Domain of institution / organisation *

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

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

[3] Your name:

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

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

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

[5] Your role in the institution:

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

[6] Your e-mail address:

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

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

Please choose only one of the following:

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

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

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

- Yes
- No

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

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

- Yes
- No

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

SECTION 2/5 - COLLECTION REGISTRATION SYSTEM APPROACH

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

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

(item level) descriptive metadata: the bibliographic and descriptive metadata needed to include item level records in your (online) catalogue.

digital born and digitally reproduced objects: *digital surrogates,* created as a result of converting analogue materials to digital form (digitisation), and *born digital objects*, for which there has never been and is never intended to be an analogue equivalent.

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

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

```

```

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

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

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

O not available online

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

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

[14] What software runs the CRS?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

GEOGRAPHY BASED RESOURCES						
	Landscapes	# Records				
	Archeological sites	# Sites				
	Monuments and buildings	# Records				
	Other geography based resources	# Records				
NATURAL RESOURCES						
	Natural inert specimens	# Records				
	Natural living specimens	# Records				
TIME BASED RESOURCES						
	Audio files: Music	# Records				
	Audio files: Speech & other (excl. digital audio books; incl. oral history files)	# Records				
	Digital audio books	# Records				
	Film (Cinematic)	# Records				
	Video recordings	# Records				
	Other time based resources	# Records				
DIGITAL INTERACTIVE RESOURCES [EXCLUSIVELY DIGITAL]						
	Databases (containing cultural heritage metadata)	# Records				

	Digital (3D) designs or reconstructions of objects and buildings	# Records				
	Digital art objects	# Records				
	Digital research files (incl. GIS files)	# Records				
	Games	# Records				
	Software (customised)	# Records				
	Websites (and parts of websites)	# Records				
	Other digital interactive resources	# Records				

SECTION 4/5 - IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

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

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

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

… %

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

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

In the table below:

- Choose all object types that apply, in other words: which object types have been digitally reproduced?

- Please estimate how these digitally reproduced objects are distributed in percentages amongst the different object types.

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

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

Which apply (A)	Term ([digital] objects)	Estimated % digitally reproduced (B)
ARCHIVAL RESOURCES		
▫	Archives : Government documents	
▫	Archives : Other archival records	
TEXT BASED RESOURCES		
Books		
▫	Rare printed books	
▫	Other printed books	
Serials		
▫	Newspapers	
▫	Journals	
▫	Other serials	
Other text based resources		
▫	Medieval Manuscripts	
▫	Other Manuscripts	
▫	Microforms / Microfilms	
▫	Other text based	
VISUAL (2D) RESOURCES		
▫	Drawings	
▫	Engravings / Prints	
▫	Maps and ground plans	
▫	Paintings	
▫	Photographs	
▫	Posters	
▫	Sheet music	
▫	Other visual (2D) resources	
3D MOVABLE OBJECTS		
▫	3 Dimensional works of art	
▫	Other (3D) man-made objects	
GEOGRAPHY BASED RESOURCES		
▫	Landscapes	

▫	Archeological sites	
▫	Monuments and buildings	
▫	Other geography based resources	
NATURAL RESOURCES		
▫	Natural inert specimens	
▫	Natural living specimens	
TIME BASED RESOURCES		
▫	Audio files: Music	
▫	Audio files: Speech & other	
▫	Film (cinematic)	
▫	Video recordings	
▫	Other time based resources	
▫	Other (please specify below)	

Other object types (not listed above) are:

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

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

… %

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

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

In the table below:

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

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

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

Which apply (A)	Term ([digital] objects)	Estimated % digitally reproduced (B)
ARCHIVAL RESOURCES		
▫	Archives : Government documents	
▫	Archives : Other archival records	
TEXT BASED RESOURCES		
Books		
▫	Rare printed books	
▫	Other printed books	
Serials		
▫	Newspapers	
▫	Journals	
▫	Other serials	
Other text based resources		
▫	Medieval Manuscripts	
▫	Other Manuscripts	
▫	Microforms / Microfilms	
▫	Other text based	
VISUAL (2D) RESOURCES		
▫	Drawings	
▫	Engravings / Prints	
▫	Maps and ground plans	
▫	Paintings	
▫	Photographs	
▫	Posters	
▫	Sheet music	
▫	Other visual (2D) resources	
3D MOVABLE OBJECTS		
▫	3 Dimensional works of art	
▫	Other (3D) man-made objects	
GEOGRAPHY BASED RESOURCES		

	Landscapes	
▫	Archeological sites	
▫	Monuments and buildings	
▫	Other geography based resources	
NATURAL RESOURCES		
▫	Natural inert specimens	
▫	Natural living specimens	
TIME BASED RESOURCES		
▫	Audio files: Music	
▫	Audio files: Speech & other	
▫	Film (cinematic)	
▫	Video recordings	
▫	Other time based resources	
▫	Other (please specify below)	

Other object types (not listed above) are:

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

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

Please choose only one of the following:

- Yes
- No
- Do not know

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

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

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

Which apply (A)	Term ([digital] objects)
ARCHIVAL RESOURCES	
▫	Archives : Government documents
▫	Archives : Other archival records
TEXT BASED RESOURCES	
▫	Electronic books (eBooks)
▫	e-Newspapers

▫	e-Journals
▫	Other e-Serials
▫	Other digital born text documents
VISUAL (2D) RESOURCES	
▫	Digital maps and ground plans
▫	Photographs
TIME BASED RESOURCES	
▫	Audio files: Music
▫	Audio files: Speech & other
▫	Film
▫	Video recordings
DIGITAL INTERACTIVE RESOURCES	
▫	Databases (containing cultural heritage metadata)
▫	Digital (3D) designs or reconstructions of objects and buildings
▫	Digital art objects (incl. Internet art)
▫	Digital research files (incl. GIS files)
▫	Games
▫	Software (customised)
▫	Websites (incl. blogs, tweets, widgets, wikis)
▫	Other (please specify below)

Other object types (not listed above) are:

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

SECTION 5/5 - EVALUATIVE REMARKS

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

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

APPROACH	# Minutes
COLLECTION REGISTRATION SYSTEM APPROACH	
DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH	
IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH	

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

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

APPROACH	1	2	3	4	5
COLLECTION REGISTRATION SYSTEM APPROACH					
DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH					
IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH					

(Using a 5-points scale - where 0 equals "very difficult/hard" to 4 "not at all difficult/very easy" - please select only one number per row. Again, please try to account for the reuse of information, as in the previous question.)

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

- COLLECTION REGISTRATION SYSTEM APPROACH
- DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH
- IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

(Please select only one of the alternative approaches.)

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

(Please explain you answer to question 25a.)

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

- COLLECTION REGISTRATION SYSTEM APPROACH
- DETAILED MEASURING OF DIGITAL SUB COLLECTIONS APPROACH
- IMPROVED ENUMERATE CORE SURVEY QUESTIONS APPROACH

(Please select only one of the alternative approaches.)

[26b] And why?

```
[blank answer box]
```

(Please explain you answer to question 26a.)

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

```
[blank answer box]
```

Thank you for collaborating with us in this Survey!

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

APPENDIX

Definitions for the survey on the SIZE of digital heritage collections

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. [GJN]**
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	**An article is a written work published in a print or electronic medium. It may be for the purpose of propagating the news, research results, academic analysis or debate.**
Audio book/Digital audio book	**A (digital) audiobook is a (digital) recording of a text being read, e.g. on CD or DVD, or by means of other (digital) technologies.**
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.**
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 for (part of) the collections of the institution are included in its mission. Institutions that have collections of heritage materials (like for example of books, films, music, etc.) to be lend by 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, games, etc.
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.
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.
Games	Forms of competitive play, usually involving an element of strategy, especially to interfere with an opponent's play.
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.
Ground plan	Floor plans taken near ground level and showing the foundation and ground floor of the building.
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.
Landscape	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.
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.

Metadata	Literally, "data about data." Structured information describing information resources/objects for a variety of purposes....The term is generally used in the library community for ⌣on-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.
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.
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. The sensitized surface may be some sort of charge-coupled device, resulting in binary signals to be electronically stored in the form of encoded picture elements, or pixels (digital photographs).
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.
Print	Copy of an image transferred to a sensitive material.
Rare printed book	Book published before 1800.
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.
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.
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.
Visual resources	Any items that communicate information primarily visually and are collected and made available as information resources.

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.

8.8 Questionnaire on the COST of digital heritage collections

Summary

This questionnaire was developed by Robert Gillesse and Gerhard Jan Nauta (both working at DEN). The starting point consisted of a few questions focussing on the cost of digital collections from the ENUMERATE Core Survey (2012). Indications and suggestions from the Specialist Meeting on Cost (London, March 30) are taken on board as well, and several surveys in the field (e.g. the recent ITHAKA S+R Survey on Sustaining Digitized Special Collections) were examined. The central part in this survey, however, is a listing of activity related costs that were deduced from research done at the DEN Foundation (2009).

An online test version of the questionnaire on COST, was developed by Panteia and links to the questionnaire were sent to about 15 institutions in three EU member states (5 institutions per member state). This is Part A of the substudy. Based upon the outcome of the testing phase, the online questionnaire will be adapted. In Part B a larger test will be done on a sample of about 100 institutions. The National Coordinators in 7 different EU member states have selected the institutions participating in Part B (10 to 15 institutions per member state).

COST

Document history

Date	Author	Status
5-2-2013	Robert Gillesse (DEN), Gerhard Jan Nauta (DEN)	Draft #1

ENUMERATE SURVEY ON THE <u>COST</u> OF DIGITAL COLLECTIONS

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

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

The survey consists of 22 questions. Since the questionnaire aims to improve the methodology of monitoring cost, several questions have an accompanying field for evaluative remarks. Please help us with any comments that may be useful.

The time needed to do the survey will depend on the availability of management information about digitisation activities. If this information is in place, we estimate that answering the questions will take about 30-60 minutes of your time. If the information is not readily available additional research in your institution may be needed, but the reward for such efforts will go both to your institution and the community of museums, libraries and archives.

We hope you will be able to complete the questionnaire by February 28. The questionnaire is designed in such a way that you can pause and return at any time to continue. After you have finished and submitted the questionnaire, we may possibly contact you in order to resolve ambiguities.

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

Natasha Stroeker: enumerate@panteia.nl

Best Regards, the ENUMERATE Team.

=== questionnaire ===

SECTION 1/3 - ORGANISATIONAL INFORMATION / INSTITUTIONAL PERSPECTIVE

[1] Name of institution / organisation *:

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

[2] Type / Domain of institution / organisation *

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

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

[3] Your name:

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

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

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

[5] Your role in the institution:

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

[6] Your e-mail address:

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

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

Please choose only one of the following:

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

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

SECTION 2/3 - THE OVERALL COST AND FUNDING OF DIGITAL HERITAGE COLLECTIONS

[8] Please specify (or estimate) the annual budget - *in Euros* - for creating and/or acquiring *new* digital heritage collections (i.e. the above mentioned <u>incidental costs</u> *) in two consecutive years.

Please specify the budget concerned in the last (closed) budget year:

Please specify the year concerned:

Please specify the budget concerned in the next (consecutive) budget year:

Evaluative remarks to this question:

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

[NOTE: In the TEST VERSION the evaluative remarks field should be often repeated (every question). In the final version these may be reduced to a few spots in the questionnaire.]

[9] Please specify (or estimate) the annual budget - *in Euros* - for providing access and managing your digital [heritage] collections (i.e. the above mentioned <u>structural costs</u> *) in two consecutive years.

Please specify the budget concerned in the last (closed) budget year:

Please specify the year concerned:

Please specify the budget concerned in the next (consecutive) budget year:

(This sum is the amount identified in your own institution's annual budget to provide for its digital collections. Please attempt to include the cost of the staff time devoted to digital collections in this estimate, but exclude external funding that is not included in your fixed annual budget. If budget year does not coincide with the calendar, please choose the calendar year that fits best (in terms of the number of months).

Evaluative remarks to this question:

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

[10] The annual sum total expended on the digital collections of your institution - Questions 7 & 8 - may be split up over in-house, out-sourced, and partnership activities. Please estimate the percentages of these activities in the last (closed) budget year:

[Present the SUM of the first question under questions [8] and [9]]

	Estimated percentage [sum adds up to 100%]
In-house activities	
Out-sourced activities	
Partnership activities (in-house/out-sourced)	
Other (please specify below)	

Other activities are:

(Free text description of digital collections related activities other than in-house / out-sourced / partnership.)

Evaluative remarks to this question:

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

[11] From what sources are your digital heritage collections funded?

Please consider the combined budget for incidental and structural costs and choose all that apply:

- Internal budgets
- Customer/client payments (online shop, tickets)
- Commercial trading (B2B)
- Public grants/subsidy
- Private investment
- Public/private partnership
- Other (please specify below)

Other sources of funding are:

(Indicate all the sources from which your digital collection activities receive funds. Grants, subsidies and investments that you include in your answer should be directly intended for digitisation efforts. All other (general) grants, subsidies and investments become part of your institution's budget.)

Evaluative remarks to this question:

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

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

- Yes
- No
- Do not know

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

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

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

Evaluative remarks to this question:

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

SECTION 3/3 - THE COST OF ONE SPECIFIC DIGITAL HERITAGE COLLECTION

In this section the focus will be on one specific major digital heritage collection within your institution. Preferably this is a prominent digital collection that has been publicly available (online) for at least one year. The collection may contain one or more types of digital objects (e.g. digitised paintings, digitised rare books, digital 3D reconstructions of buildings, digital photographs, etc.) and will have the associated metadata recorded in a collection registration system. (The collection registration system may also contain metadata records without associated digital objects, though.) If your institution does not have a digital heritage collection exactly matching the description above, please select the collection that comes most close.

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

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

In order to understand the digital collection at hand, we ask you to indicate what digital objects it contains, in what quantities, and in what qualities. The quality attribute can only be an approximation. Choose the highest quality that applies. We distinguish roughly:

- **High quality**: preservation or authentic copy - that is: highly faithful reproduction of the essential characteristics (e.g. contrast, colour, detail) - of the analogue original.

- **Medium quality**: reproduction that is a faithful, but not necessary 100% authentic copy of the analogue original. One or more essential characteristics might be compromised (e.g. lessened colour faithfulness).

- **Low quality**: access copy that has a loose relation to the analogue original. This kind of digitized object often has one characteristic of the original that is focused on. For instance a bitonal scan of book (high contrast image purely meant for reading).

[16] What quantities of what digital object types are in the selected collection and in what qualities?

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

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

	Term ([digital] objects)	Estimated # of units in entire digital collection	Units (#=Number of)	Quality digital/digitised object [in three levels]
ARCHIVAL RESOURCES				HQ / MQ / LQ
	Government documents		# Records	
	Government archives		# Archives	
	Other archival records		# Records	
	Other archives		# Archives	
TEXT BASED RESOURCES				
	Books			
	Rare printed books		# Records	
	Other printed books		# Records	
	Electronic books (eBooks)		# Records	
	Serials			
	Newspapers		# Issues	
	Journals		# Issues	
	Articles		# Records	
	Other serials		# Issues	

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

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

Evaluative remarks to this question:

```

```

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

[17] What are the file formats of the digital objects (e.g. TIFF, JPEG, PDF, PDF/A, SVG etc).

```

```

(Please fill one or more file formats, separated by commas.)

Evaluative remarks to this question:

```

```

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

[18] Does the collection have any copyright restrictions?

- Yes
- No
- Do not know

[19] If Yes, approximately which part (%) of the collection has been cleared of rights/licensed for giving online access?

```

```

Evaluative remarks to this question:

```

```

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

[20] Please estimate the percentage (%) of the digital collection that is available online:

```

```

Evaluative remarks to this question:

```

```

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

[21] Activity based costing: determining Incidental costs

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

[The listing of cost items is based on a cost model developed by the DEN Foundation. In the test version of this survey the aim is to get reliable input for a breakdown of costs. On the basis of the feedback a new, more thorough overview of costs will be constructed.]

(Please note that the time period concerned can be indicated below the table.)

[NOTE: We can probably hide the "Total #EURO"-column.]

[NOTE: For the Other costs at the bottom of the table we need repeatable rows.]

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

(In the #EURO cells enter a whole number without a currency symbol or commas. Please use the Remarks field to comment on the cost item concerned: is it familiar? Can it easily be determined? Is it composed of sub-items?)

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

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

Evaluative remarks to this question:

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

[22] Activity based costing: determining Structural costs

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

For *the past fiscal year*, please estimate the structural costs for the selected digital collection. Fill in the form below for each of the following activities, wherever it occurs in the institution or is outsourced. Please enter a whole number without a currency symbol or commas.

 (In the #EURO cells enter a whole number without a currency symbol or commas. Please use the Remarks field to comment on the cost item concerned: Is it familiar? Can it easily be determined? Is it composed of sub-items?)

Evaluative remarks to this question:

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

[TESTVERSION:]

Once the ENUMERATE Team has evaluated your answers, we may possibly seek contact with you by telephone in order to resolve ambiguities and/or do a walkthrough with some of the questions. Please let us know if you agree with that.

Yes, I agree with feedback by telephone.

No, I don't want to be contacted anymore.

[/TESTVERSION]

Thank you for collaborating with us in this Survey!

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

APPENDIX

Definitions for the survey on the COST of digital heritage collections

Annual revenue budget	The annual budget for the entire institution, for instance as indicated in the last set of published accounts. The budget may include government funding, project funding, revenues from the organisation's primary or commercial activities, etc.
Archival resources	An artificial accumulation of materials devoted to a single theme, person, event, or type of document acquired from a variety of sources.
Collection	Gathering of documents assembled on the basis of some common characteristic, without regard to their provenance.
Cultural heritage domain	The 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 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 (like for example of books, films, music, etc.) to be lend by or sold to contemporary users without the explicit task of safeguarding the collections for future generations, are not included. This essentially leaves out both school libraries and public libraries without cultural heritage collections.
Digital interactive resources	Products and services on digital computer-based systems which respond to the user's actions by presenting content such as text, graphics, animation, video, audio, games, etc.
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.
Geography based resources	Geographic information is any data with direct or indirect reference to specific location or geographical area on the surface of the Earth and can be in the form of geographic coordinates and/or geographic names.
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.
Metadata	Literally, "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.
Structural cost (of digital collections)	The sum total of the 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.
Time based resources	Includes cultural heritage objects that are not static but evolve during a certain time frame like video, audio, film, etc.

8.9 Questionnaire on the USE of digital heritage collections

Summary

This questionnaire on the use of and access to digital heritage collections was developed by Wietske van den Heuvel and Gerhard Jan Nauta (both working at DEN). Again the starting point consisted of questions focussing on, in this case, the use of digital collections from the ENUMERATE Core Survey (2012). These questions were amplified by a number of questions that were developed with suggestions from the Specialist Meeting on Use/Access (Madrid, 6th June 2012) in mind. Other sources are recent surveys in the field, e.g. the ITHAKA S+R Survey on Sustaining Digitized Special Collections and especially small survey research that was conducted in the Netherlands by Henk Voorbij (2009).

An online test version of the questionnaire on USE/ACCESS, was developed by Panteia and links to the questionnaire were distributed among about 30 institutions in 5 EU member states.

USE

Document history

Date	Author	Status
5-2-2013	Wietske van den Heuvel (DEN), Gerhard Jan Nauta (DEN)	Draft #1

ENUMERATE SURVEY ON THE <u>USE</u> OF DIGITAL COLLECTIONS

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

Online access: which part of your collection is available online and what are the platforms where content is presented to the public?

The development of online user services: has your institution developed any websites, apps, etc.?

The measurement of online user activities: what methods and tools does your institution use for collecting usage metrics?

Licensing and legal issues regarding the online collection: how much does your institution spend on clearing copyright?

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

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

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

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

Natasha Stroeker: enumerate@panteia.nl

Best Regards, the ENUMERATE Team

=== questionnaire ===

SECTION 1/6 - ORGANISATIONAL INFORMATION

[1] Name of institution / organisation *:

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

[2] Type / Domain of institution / organisation *

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

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

[3] Your name:

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

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

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

[5] Your role in the institution:

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

[6] Your e-mail address:

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

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

Please choose only one of the following:

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

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

SECTION 2/6 - POLICY

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

- Yes
- No
- Do not know

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

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

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

O not available online

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

- Yes
- No
- Do not know

Evaluative remarks to this question:

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

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

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

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

Other types of use are:

Evaluative remarks to this question:

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

SECTION 3/6 ACCESS TO ONLINE DIGITAL COLLECTIONS

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

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

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

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

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

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

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

	resources		
	Medieval Manuscripts		# Records
	Other Manuscripts		# Records
	Microforms / Microfilms		# Records
	Other text based		# Records
VISUAL (2D) RESOURCES			
	Drawings		# Records
	Engravings / Prints		# Records
	Maps and ground plans		# Records
	Paintings		# Records
	Photographs		# Records
	Posters		# Records
	Sheet music		# Records
	Other visual resources		# Records
3D MOVABLE OBJECTS			
	3 Dimensional works of art		# Records
	Archeaological man made objects		# Records
	Coins and medals		# Records
	Other 3 dimensional man-made objects		# Records
GEOGRAPHY BASED RESOURCES			
	Landscapes		# Records
	Archeological sites		# Sites
	Monuments and buildings		# Records
	Other geography based resources		# Records
NATURAL RESOURCES			
	Natural inert specimens		# Records
	Natural living specimens		# Records
TIME BASED RESOURCES			
	Audio files: Music		# Records

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

Evaluative remarks to this question:

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

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

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

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

Other kinds of metadata are:

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

Evaluative remarks to this question:

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

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

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

Other access options are:

Evaluative remarks to this question:

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

SECTION 4/6 - CONDITIONS OF USE

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

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

Evaluative remarks to this question:

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

SECTION 5/6 - MONITORING

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

Please choose only one of the following:

- Yes
- No
- Do not know

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

Evaluative remarks to this question:

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

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

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

Please choose all that apply:

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

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

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

Evaluative remarks to this question:

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

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

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

Monitoring tools are:

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

Evaluative remarks to this question:

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

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

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

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

Evaluative remarks to this question:

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

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

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

The analysis of access/use data is used for:

```
[                                                                      ]
```

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

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

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

- Yes
- No
- Do not know

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

```
[                                                                      ]
```

(Please briefly mention the KPÍ's that are used within your organisation.)

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

```
[                                                                      ]
```

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

Evaluative remarks to this question:

```
[                                                                      ]
```

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

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

```
[                                                                      ]
```

SECTION 6/6 - FOLLOW-UP SURVEY

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

- Yes
- No

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

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

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

[25] E-mail address:

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

QUESTIONNAIRE EVALUATION

[26] How long did it take to fill out the questionnaire?

[27] Did you have to look up things before filling out certain questions?

Thank you for collaborating with us in this Survey!

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

APPENDIX

 Definitions for the survey on the ACCESS TO/USE of digital heritage collections

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.
Collection	Gathering of documents assembled on the basis of some common characteristic, without regard to their provenance.
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 for (part of) the collections of the institution are included in its mission. Institutions that have collections of heritage materials (like for example of books, films, music, etc.) to be lend by 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.**
Digital collections	**The whole of digitised and born-digital cultural objects and their metadata within a cultural institution.**
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.**
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.**
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.**
Geoblocking	**Restricting access only for users from a certain geographical region (e.g. a country).**
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.**
Item-level metadata	**The bibliographic and descriptive metadata needed to include item level records in your online catalogue.**
Key Performance Indicators	**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.**
Metadata	**Literally, "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.**
National aggregator	**A website that aggregates the digital collections of a nation's heritage institutions and serves as an access point to these collections.**
Offline	**A digital collection that is available on the local area network of the institution and which cannot be accessed through the internet.**
Platform	**A (computing) platform includes a 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).**

Remix	**In the context of the ENUMERATE 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.**
Thematic aggregator	**A website that aggregates digital collections that are thematically linked and which serves as an access point to these collections.**
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.**
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.**
Web analytics	**Is the measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing web usage.**