



Developing a statistical framework for measuring the progress made in the digitisation of cultural materials and content.

A summarised version of the study report on:

Desk-Research Findings

A summary of the research undertaken in the first phase of the study to take account of relevant work in the same field and scope the target population for measuring digitisation activities relating to Europe's cultural heritage.

The opinions expressed in this study are those of the authors and do not necessarily reflect the views of the European Commission.

IPF Ltd
No 1 Croydon
CR0 0XT
United Kingdom

IPF

November, 2007

<http://www.NUMERIC.ws>

Desk-Research Findings

Contents

Section		Page
1	Objectives	1
2	Research Approach	1
3	Literature Review	3
4	Consultation Feed-back	11
5	Foundation Database Sources	15
6	Framework Design:	18
7	Review of Classifications / Definitions	20
8	Definition of the Universe	27

Appendix		
I	Bibliography	30

Digitisation is allowing, so called, "memory institutions" to diversify and enrich the information economy with Europe's cultural heritage. The Numeric study is establishing a framework to measure the pace and cost of translating often frail collections into this rich resource.



Reliable measures (statistics) will better inform the way in which the digitisation process is managed:

- * Just how many institutions are digitising heritage materials?
- * How much has been spent, and how much more is needed?
- * What proportion of collections, can be read on the internet?

There are so many more questions policy makers are asking, and managers responsible for the work are interested in measuring their own performance.

1 OBJECTIVES

1.2 The European Commission¹ has adopted a recommendation to “gather information about current and planned digitisation of cultural material”. The Numeric study aims to establish a statistical framework to implement this recommendation and, at the outset, the study team have conducted wide-ranging research to consider the most practicable approach. This desk-research phase has lasted six months and worked to the following requirements:-

- * To carry out thorough desk research to identify all relevant data sources, acquire and compile a first dataset and statistics, and identify indicators and methodologies used.
- * To identify and define the target population in the domains of libraries (covering at least national, local, and research libraries), archives (including audio-visual archives), and museums (also including public galleries and natural history museums). This will inform the sample frame for each country, for the collection of data as the study progresses.
- * To follow as much as possible the definitions of cultural institutions as used by UNESCO and EUROSTAT.

1.3 The following report, therefore, presents the results of this research, including: existing relevant data, sources, statistics, indicators and the concepts, definitions and classifications that should apply for determining the statistical framework that can measure the extent and cost of cultural heritage digitisation in Europe.

2 RESEARCH APPROACH

2.1 During the inception phase of the study it was also agreed to undertake a number of parallel actions to underpin the development of the statistical framework. These included:

- * mobilising and engaging stakeholders by seeking representation of experts in each country and conferring with the professional bodies active at international level;
- * raising awareness amongst policy makers who are influential in funding digitisation programmes, by presenting the objectives of the study and informing them of developments by newsletter (and web-site);
- * assembling a “foundation” data-base comprising statistics and inferred estimates gained from the Culture Statistics Observatory and other country profiles contained in such

¹ Recommendation 1, on the digitisation and on-line accessibility of cultural material and digital preservation, of 24 August, 2006.

sources as EGMUS, The European Institute for Comparative Cultural Research (ERICarts), OCLC and UNESCO. The aim is to set out the number of cultural institutions by sector (Libraries, Museums, Audio-Visual Collections and Archives) and supporting financial information and data relating to analogue collections by type;

- * undertaking a canvass of all stakeholders (invited by newsletter) to reference any useful data sources;
- * reviewing the approaches and findings of prior and current studies, in order to inform the design of the NUMERIC framework by the most successful methods, definitions and classifications.

2.2 Accordingly, the above activities were initiated and have helped to add context to the desk-research findings. However, it became immediately apparent that there were few comprehensive studies on digitisation capable of providing a statistical base upon which to base reliable estimates² and it was decided to undertake a "pathfinder" survey to both gather data and test definitions and classifications.

2.3 The approach taken to these scoping research tasks, and agreed on acceptance of the study inception report, can be summarised under three main headings:

- * REVIEWING LITERATURE - taking note of the experience and understanding built-up in previous research and deriving from other associated projects in order to inform a uniform framework and consistent definitions for measuring digitisation activity – i.e. getting the definitions and classifications right;
- * ESTABLISHING RELEVANT DATA SOURCES - building a better understanding of the universe within which the digitisation activity is taking place – i.e. establishing the foundation database for scaling statistical estimates;
- * INVESTIGATING INDUSTRY SUPPORT - seeking the assistance of industry experts in locating and costing digitisation projects and reviewing the possibilities for the use of persistent identifiers for robotic approaches to collecting data.

2.4 This phase of the study was conducted in June to November, 2007.

² Only two surveys relating directly or partially to digitisation activities were being conducted in Europe: An EDL project survey of National Libraries relating to the single digital collection, coordinated by the Austrian National Library for the Conference of European National Librarians, and a survey undertaken by the European Group of Museums (EGMUS) which seeks information on the proportion of analogue collections that have been digitised – the results of the latter will not be available until 2008.

3 REVIEW OF LITERATURE

3.1 A study was made of the work of other researchers and an analysis of previous attempts to measure the progress of digitising cultural materials was conducted. This review is reported in more detail in a separate report³. A total of 32 reports selected for their particular relevance to digitisation of the cultural heritage were analysed to explore general trends, useful experiences and pitfalls in measuring digitisation in Europe and worldwide. These reports are listed in appendix 1. Several points emerged from the review that serve to illustrate how the design of the monitoring framework needs to be improved:-

- * Previous digitisation surveys have been mostly snapshots with no attempt to establish a continuous series of measures capable of describing either trends or progress from a baseline;
- * These studies have attempted to trace the perspective from the top-down (policies that relate to institutions in common) and bottom-up (management information needs for the institutions). Both focus on the need for resources, but both have failed to define consistent measures for either financial or technical resources;
- * Digitisation surveys have concentrated on collecting qualitative rather than quantitative information, and consequently few hard measures from which policies can be better informed have emerged. Consistent definitions are required for measures that indicate relative priority and value for policy determination;
- * There are no uniform approaches on how to classify digitised collections. Few studies have successfully been able to relate the way in which digitised outputs can be compared with the analogue collections. Often outputs are recorded as digital files, images, pages or collections, making meaningful analysis impossible. Setting common input/output guidelines is the single most important requirement in establishing a successful statistical framework. The units for measuring these differ for each domain and type of material in collections;
- * Project objectives reveal different levels of aspiration in the digitisation work being pursued. The basic level would suggest that an institution ignores the social or cultural benefits of digitisation beyond its own organisation; an intermediary level would refer to digitisation for online access

³ *“An analysis of state-of-the-art in measuring the progress of digitisation of cultural materials.”*
http://www.numeric.ws/download.asp?filename=NUMERIC_Desktop_Research_on_Digitisation_Studies26102007114936.pdf

and preservation of the materials; and the highest level of development sets out to exploit the digitised collections by developing educational, tourism, research and other such benefits;

- * Indicators of user needs, usage of materials, accessibility and usability of digitised content should be more visible in digitisation surveys and treated as important components of digitisation progress.

- 3.2 In addition to these studies, the design of surveys undertaken by the Conference of European National Librarians (CENL) and the European Group of Museums (EGMUS) were considered. These surveys currently attempt to relate the digitised output to the size of analogue collections held by their member institutions, and both have provided valuable insights into the difficulty of gaining such measures in a consistent manner.
- 3.3 Regard was also taken of the enormous volume of research previously undertaken and currently being pursued by the MINERVA and MICHAEL projects. Whilst the information collected provides useful context, particularly in regard to recording specific initiatives and significant digitisation projects, the summaries do not provide a base upon which consistent statistics can be collated or reliable estimates derived.
- 3.4 Taking these conclusions together, the main findings from the desk-research fall under various headings:

Defining the target for surveys

- 3.5 The target of surveys in the reports we reviewed inevitably varied according to the geographical coverage of the study:
- * International – reports that cover respondents from several countries were usually performed by the international professional organisation, international consortia or projects.
 - * National – reports that cover diverse national institutions were commissioned by national governments and/or major Film / Archive / Library / Museum professional bodies.
 - * Local – reports that cover a small sample of institutions (in one country), were developed for local needs of smaller professional bodies.
- 3.6 Usually the studies targeted their surveys at specific institutions that were assumed to perform digitisation activities (13 reports, 41% list 'specific type of institution' as a criterion for establishing target population). Often these are archives, libraries and museums but also governmental agencies, representatives of the commercial sector etc. The second widely used criterion for establishing the target population is membership of a professional

organisation or consortium (8 reports, 25 %). In many cases a digitisation survey or any other related research was initiated by the association interested in the current state-of-the-art amongst their member institutions.

- 3.7 Often a target population was defined but the sample was not representative and the less than full response possibly distorted the results even further. **The one factor common to all the studies we reviewed was the difficulty in gaining comprehensive responses to sample quotas, and this is likely to be the most significant risk to the NUMERIC framework, once it is implemented. To rely on surveys alone is not an option.**
- 3.8 However, several reports demonstrated useful sampling designs that allow more precise data about digitisation activities to be obtained. For instance, in Digitisation and the Preservation of Digital Data in German Museums (Germany, 2005) and Digitisation and preservation in public libraries and archives (UK, 2002) several criteria for defining a sample, such as level of computerisation of a museum in the first case and an analysis of previous research and directories of digitisation projects in public libraries and archives in the second case, were applied. The survey Status of Technology and Digitisation in the Nation's Museums and Libraries applied a stratified sampling technique to create a controllable sample of US archives, libraries, museums and library agencies (US, 2006).
- 3.9 In order to determine what institutions were covered and how they were represented in the reports, an analysis of responders was performed. The majority (114 of 149) institutions mentioned are museums (53), libraries (42) and archives (19). Further analysis reveals that each type of memory institution is represented in different ways. Libraries are presented in the most structured manner, using widely accepted typologies (e.g. academic libraries, public libraries, national libraries etc.). Analysis of the categories describing museums reveals that there is still no widely accepted typology of museums that would be employed for research purposes, excepting that adopted by EGMUS. The "Archive" is the most undifferentiated domain in comparison with libraries and museums. Often researchers refer to these institutions in an abstract manner ('archives'), in some cases they are incorporated as part of a library or museum. Similarly the institution holding audio-visual collections can be an "Archive" (e.g. Radio Archive) Broadcaster, film institute, or sound museum.
- 3.10 However, due to increased attention at the international level to the audiovisual heritage, it is necessary to consider a concept of audiovisual archive, though it was not sufficiently distinguished in reports we have reviewed. The definition of 'audiovisual archive' is primarily based on management of specific collections (e.g. see Edmondson, 2003). Such a definition assumes that any institution

performing the functions described should be considered an audiovisual archive – e.g. museum, library or archive. In spite of the fact that traditionally archives, libraries and museums manage audiovisual collections and have similar social responsibilities, there are differences in management philosophies. From a practical point of view, **the lesson we learn from this is to specify sample frames that are not too dogmatically drawn on precisely defined domains**, but reflect cross-domain institutions that naturally occur in each country. This means that the surveys must be capable of providing for all types of collection, in multi-purpose institutions.

Satisfying “Top-down” and “Bottom-up” objectives

- 3.11 The top-down perception reflects the information requirements of the responsible cultural minister or other governmental body and is directed towards measuring achievements of memory institutions in digitisation by comparing input (resources invested in digitisation) and output (results of digitisation activities). A bottom-up approach reflects the management needs of the memory institution to measure their progress in digitisation. The desire is for in-depth information about the product created by investing certain development resources and techniques (enabling its value to be judged). We were anxious to see whether these two view points could be reconciled in order to design a framework that is both practicable for individual managers from the source data will derive, and the strategic needs of policy makers faced seeking the information to justify difficult Resourcing decisions.
- 3.12 Our research shows that the top-down perception was taken in ten of the 32 analysed reports (31%). These were ministries and other governmental bodies, including agencies or departments responsible for cultural heritage or libraries, museums and archives. Twenty two reports (69%) were initiated by professional organisations, research or higher education institutions, libraries and consortia. In most cases digitisation research was initiated by the appropriate professional body (28% of all reports).
- 3.13 Top-down reports were focused on topics such as perceived obstacles, partnership in digitisation, general data on digitisation, digitisation policy, budgeting, and audiences. Bottom-up reports drew more attention to: how to select materials for digitisation, search tools / facilities / parameters, outsourcing of digitisation services, metadata, digital asset management, systems, staff issues, storage and costs. Dominant themes in top-down reports relate to strategic planning decisions while bottom-up reports mostly refer to the implementation stage of digitisation initiatives.
- 3.14 The top-down approach overlooked methodological solutions and proper management of resources that affects the quality of the

digitisation products; while the bottom-up approach largely ignored the longer-term strategic objectives that are also indicators of digitisation progress; (possibly because of uncertainty relating to the availability of resources). **In order to reconcile these perspectives and seek common priorities we consulted stakeholders directly**, and have reported the conclusions below.

Accounting for the cost / funding issues

- 3.15 Most of the studies we reviewed were concerned with tracing funding for digitisation initiatives (11 reports, 34%). For these the major question was about the source of funding, including state funding (local and national), institutional budgets, corporate funding, foundation grants, gifts and other revenue sources. Exploitation of digital services (incl. those based on digitised collections) appeared to be a controversial issue, set beside the principle of memory institutions deriving funds from the tax-payers. The issue could equally be covered around the topic of copyright. However, the focus of more recent studies has identified an opportunity for charging for digital cultural heritage services (see DigiCULT report *Technological landscapes for tomorrow's cultural economy: unlocking value of cultural heritage*, 2002).
- 3.16 In five reports (16%) respondents were asked to quantify their total digitisation expenditures: Annual digitisation budget/expenditures (three reports) and sum of all digitisation expenditures to-date (two reports). In the analysis of digitisation activities in academic and research libraries, "Managing digitisation activities" (2006, US), budget allocations for 'start-up' and 'ongoing operations' were distinguished. A distinction was made between spending when an institution begins digitisation activities and builds technological infrastructure (start-up), and further spending when the infrastructure is in place (Mugridge, 2006). **For our own purposes, particularly in collecting information from undeveloped accounting bases, such a distinction seems to unnecessarily confuse the primary objective of gaining consistent financial data.**
- 3.17 Money spent on digitisation at the state level (i.e. programmes providing certain amount of funds) may indicate a strategic importance of digitisation at the national level. Two reports – Digitised content in the UK research library and archives sector (UK, 2002) and a report on digitisation activities in Lithuania performed by the Ministry of Culture in collaboration with Vilnius University (LT, 2006) demonstrated an attempt to summarize financial data on the national digitisation spending. However, researchers (Bültmann, Hardy, Muir and Victor, 2006) often report problems obtaining precise data on digitisation funding at the national level, and **there is no evidence available from the studies we have**

reviewed to prove that the memory institutions account for the activity in a uniform manner.

- 3.18 None of the studies proposed any guidelines for calculating costs. Hence, the unspecific survey questions are possible reasons for low response. For instance, in the survey Digitisation and preservation in public libraries and archives (UK, 2002) respondents were asked to provide costs of maintenance of hardware and software, but the responses are ambiguous as to whether staff time devoted to maintenance procedures have been included. (Astle and Muir, 2002).
- 3.19 Digitisation costs have been discussed in the international literature; and recommendations based on project experiences or drawn by expert teams have provided various methods for making the calculations. These include a handbook, "Assessing the costs of conversion: Making America IV: the American Voice 1850-1876" (2001), a widely cited article Digitisation: "How much does it really cost?" (Tanner and Smith, 1999), "RLG worksheet for estimating digital reformatting costs" (RLG, 1998), "Guidelines for digitization projects for collections and holdings in the public domain", particularly those held by libraries and archives (UNESCO, 2002), the Prestospace cost calculator⁴, and the British Library Cost Model⁵.
- 3.20 From these approaches it is possible to distinguish five cost factors: digitisation objectives, collection parameters, quality of digital surrogate, application of cost reduction mechanisms and technological infrastructure. Cost factors affect certain stages of digitisation. Notably, all recommendations cover materials that are usually held in libraries and archives. At the moment there are no in-depth studies covering methodology for calculation of digitization costs for museum objects.
- 3.21 Despite the complexity of the digitisation procedures, we recognise that **a much simpler standard form for accounting for digitisation activities is required, to provide a more digestible guide for institutions that are asked to report such costs.**

Input – Output Measures

- 3.22 Just under one-half (15 or 47%) of the reports we reviewed provided various definitions for composites of 36 different types of digitised materials. Of all the types, the term 'audio-visual' is the most ambiguous. Definitions of audio-visual labels vary highly from those encompassing still images to narrower definitions concentrating on audio/video/film materials alone. For instance, in UNESCO's Survey of endangered audio-visual carriers, posters and

⁴ see <http://www.prestospace.org/>

⁵ "Costing the Digital Preservation Lifecycle", Paul Wheatley, Digital Preservation Manager, BL..

still photographs are covered (Boston, 2003). The CLIR report "Evidence in Hands" describes still images, including maps, photographs, posters, cartoons and other similar materials as audio-visual content (Evidence in hand..., 2001). On the other hand, in the surveys undertaken by the TAPE and PRESTOSPACE projects, only audio and video materials were covered. In definitions of audio-visual documents, the need for special equipment to transmit or present the material is emphasized (Edmondson, 2003). In this research, only film, audio and video recordings are considered to be audio-visual materials.

- 3.23 Five studies provided classifications by "subject". In these reports respondents were asked to indicate thematic coverage of the collection (e.g. arts, humanities, religion, science etc.). In a questionnaire developed by the EDL project (see <http://edlproject.eu/>) institutions were asked to provide not only subject information but also geographical and temporal characteristics of the collection.
- 3.24 Ten (31%) reports sought to express "achieved" or "planned" digitised outputs as relative measures of the analogue collection. Three measures were adopted: 1) number of analogue materials that are planned to be digitised; 2) proportion of the analogue collection already digitised; 3) materials in analogue collection that are suitable for digitisation. The first two are relatively easy to conceive, although more difficult to express in comparable input / output terms. The third measure is both ambiguous (what criteria are used to define "suitability"?) and impossible for a single institution to report in a consistent manner with any other – i.e. the priorities of both may be very different, and each would need to work from the same "index of need to digitise": How would conservation / preservation priorities rate against the desire to widen access to the materials?
- 3.25 The estimates of the proportion of analogue collections translated into digital form, in many cases, have to be interpreted as approximations. Often analogue measures are used to make the transformation comparison (e.g. physical items in the collection); in some cases the measures were digitised images – but then the analogue half of the equation has not been expressed. **The survey undertaken by the CENL attempts to address this difficulty by applying an average number of pages to volumes in the analogue book collections, and this approach may lend itself to approximating the input / output relationship for other types of material, based on other appropriate units.**
- 3.26 The literature provided no clear standard for measuring digitised materials. Measures of digitised outputs such as the 'number of images/files/pages' reflect the technological process of conversion. However, such measures on their own reflect different standards,

such that one TIFF file is obviously quite different to another that has also progressed to OCR.

Drivers of Digitisation Projects

- 3.27 Eighteen of the studies we reviewed (56%) explored the general objectives or motivation of institutions (reasons) to digitise their collections. Three main drivers were identified: 1) access (mentioned 39 times, 43%); 2) information management issues (mentioned 36 times, 40%); and 3) preservation (mentioned 16 times, 17%). This suggests that all these factors influence policies, and this was borne out in the feed-back we received during our direct consultations with institutions. In response to our own survey, responders were often unable to provide a distinction between the materials digitised for preservation purposes and those converted to widen access.
- 3.28 Widening access was pursued to enhance educational services, awareness of institutions/collections, and research opportunities. Expressed objectives included: "improve access to unique materials", "more people will be able to access the resources", and "present more of the collection than is on display at any one time". **Breaking down the statistics to justify the various drivers of digitisation programmes is over-complicated and difficult. It is clear that a general measure of overall progress will better inform policies**, that are as diverse as extending support to either teaching or learning processes, developing cultural tourism and entertainment (leisure) services.
- 3.29 Further objectives were pursued to develop institutional efficiency, such as saving space, generating revenue, informing marketing, and providing research opportunities. The benefit most often cited was the improvement to support internal resource management processes, such as documentation, expanding databases, improving document delivery, and cataloguing.
- 3.30 Three main reasons for pursuing preservation policies arose: decreasing direct access to frail or rare originals, preserving the originals, and migrating from obsolete carriers. The most frequent motivation cited was decreasing usage of the originals. In some cases, especially in audio-visual archiving, short life-cycle digitisation is a function of the limited technical life of the carriers (equipment used to read the material). This raises an interesting point concerning the **need to monitor the "up-grade" effort inherent in digitisation programmes**: Both in proceeding to interpret the text (OCR) from simple image files, and to transfer audio-visual content from obsolete to formats that offer better standards.

End-User Considerations

- 3.31 Eleven studies (34%) explored user needs, usability and accessibility to digitised outputs. Two were entirely dedicated to examining user needs: These were the IMLS report "Assessment of End-User Needs in IMLS-Funded Digitization" (2003, US) and "DLF-Aquifer Services Institutional Survey Report" (2006, US). The focus related to usage statistics, which covered the number digital service users, number of visitors, and frequency of use. Some reports were interested in general information on whether institutions conduct user needs assessment.
- 3.32 The majority of studies we reviewed did not outline a framework for the continued monitoring of digitisation activities and, to be fair, this was not their stated intention. However, **it is clear that proof of a growing audience for digitised materials is an essential indicator to support the policy debate.**
- 3.33 A more detailed analysis of the studies we reviewed is available under the title, "An analysis of the state-of-the-art in measuring the progress of digitisation of cultural materials", published on the NUMERIC website. This provides an invaluable summary of the variable success of different categorisations of materials included in cultural heritage collections. Further findings from the review inform many of the conclusions we set out later in this report.

4. CONSULTATION FEED-BACK

- 4.1 To provide some structure to our consultations, a policy pro-forma was made available on the Numeric web-site. We invited ministries, professional associations, institutions and other interested stakeholders to provide views on the sort of information the statistical framework should provide. We considered it important to seek such views, rather than depend solely on reviewing prior studies, many of which had more restricted ambitions than developing a cross-domain / pan-European framework.
- 4.2 Eligibility to qualify for a prize-draw was advertised by way of thanking responders for their consideration. The consultation framework will remain open until 1 December, when the prize draw will take place, and the comments provided by 52 responders at the time of writing this report are summarised in this section.
- 4.3 The overwhelming message provided was **the need to establish the link between the investment in digitisation projects and the benefit to those persons who would gain access to the materials.** "There's a lack of information about the access and the use of digitised content by the public; having this sort of information could really address and support the future of digitisation initiatives", was a typical response.

- 4.4 Another angle on the same point was expressed as follows: "There has been significant investment into digitisation. Unfortunately, I am still to be convinced that what has been digitised is being used to its capacity." **The importance of gaining hard facts to inform funding policy decisions was recognised by everyone.**
- 4.5 Consultees said that many issues were equally important, and ranking them was therefore impossible, but preference was shown for information about the "Cost" and "Access / Usage" of digitised materials, followed by statistics about the progress "Achieved" and "Planned". To an extent, information about the "Formats" was regarded as contextual rather than of strategic importance for the policy debate.
- 4.6 "Funding" was reported as the single most problematic issue relating to digitisation projects, "Copyright" issues were also significant. The point was also made that, once the materials had been digitised, further investment was needed to present the information to users. However, the positive impact that digitisation programmes were making in benefiting future generations and opening up access now to the cultural heritage for all was stressed.
- 4.7 The worth of developing the statistical framework is recognised – but unless institutions respond to future surveys, the statistics that are capable of influencing policy will never be generated. This really is an instance where actions count more than words! It is equally clear that unless both ministries and institutions believe that their **efforts to generate the data will really stimulate funding flows**, the actions will not happen. This will be **a key issue for all interested parties to consider at the planned workshop to consolidate the framework approach** in the Spring of 2008.

Technical Considerations

- 4.8 A more specific exercise was undertaken to test the alternative classifications and indicator definitions to be employed in the framework. This was pursued by selecting the more successful classifications and definitions observed in the desk-research and formulating a questionnaire that was then tested in a "Pathfinder" survey. The questionnaire invited feed-back relating to the design of the survey so that we could refine the definitions accordingly
- 4.9 The survey questionnaire is available from the Numeric web-site, and was launched in September 2007 with the assistance of the Ligue des Bibliothèques Européennes de Recherche (LIBER), The Consortium of European Research Libraries (CERL) and the Association des Cinémathèques Européennes (ACE). In addition, persistent professional and ministerial sponsorship was given to the exercise in Hungary, Portugal and Sweden. The survey is ongoing, and the data and more complete survey results will be made available to all sponsoring organisations. In the meantime, the

lessons learned to-date have influenced the framework approach recommended later in this report.

Financial statistics

- 4.10 The pathfinder survey was designed to collect information about costs in two ways: Directly seeking unit costs according to specific standards; requesting information about the volume of planned work and the total expected cost, from which unit costs could be computed when analysing the data. However, it is evident from the results that this proved the most problematic area for most respondents, and this was reflected in the paucity of information provided in the returned questionnaires. Comments provided also raised the need to define exactly what should be included in the cost, and in particular, whether or not it should include a recharge for staff time. **For consistency's sake, staff time should be recharged to the reported expenditure in future surveys.**
- 4.11 Furthermore, it was suggested that costing information was commercially sensitive. However, the consensus of views, shared by external contractors we have consulted, is that the **cost varies according to the standards sought, volume of work, and way in which the project is organised.** The publication of illustrative benchmarks can only serve to better inform local judgements concerning both the value and relative complexity of planned projects. Therefore, **this sort of qualitative detail will be required from the framework.**
- 4.12 Some respondents pointed to the difficulty of costing plans because the estimates often proved to be very wide of the actuality. Others stated that it was more difficult to trace the cost of work completed, because the accounting information was unavailable. However, the digitisation agenda is so substantial and the investment required so significant that it is evident that more importance needs to be given to both planning and accounting for the activity, both within the institution and amongst sponsoring ministries. **The framework needs to recognise sufficient lead time required for institutions to adopt the guidance on how to account for digitisation work.**

Input – Output measures

- 4.13 The units required to measure the analogue collections and the digital outputs need to coincide, but the definitions need to reflect more convenient ways in which the relevant institutions maintain their management information. The way in which bound printed materials can be reported in both numbers of volumes and pages is useful in deriving consistent measures. However, to minimise ambiguities in the statistics more specific guidance and measures need developing for materials such as the following:

- * Newspapers and journals – issues / titles / editions / articles;
- * All loose printed or manuscript materials – bundles / files / collections. Photographs, in particular, are often difficult to quantify as single images;
- * Museum objects – specimens (scientific) / artefacts (man-made) / works of art;
- * Distinct types of audio-visual materials – conversion of analogue materials / upgrade to new digital formats.

4.14 Accordingly, **a revised approach to matching the measures for these materials in their analogue and digitised forms is recommended** in section 7.

Programme Design

- 4.15 The concept of prioritising materials for digitisation appears to be a philosophical one. A better definition is required to identify what proportion of the analogue collection coincides with a completed digitisation programme. The most prevalent approach taken by institutions is to work their way through the digitisation of specific collections. This defines their priorities, and the available budgets (whether notional or formally earmarked) mark the speed at which these projects progress.
- 4.16 The survey design needs to recognise the stages in the digitisation process. Partial digitisation can commence with the catalogues, prior to the conversion of the materials to which these refer. This can be catered for by including **a measure covering the proportion of the catalogue available on-line** (restricted or open access).
- 4.17 Another stage in the process relates to the proportion of materials in image format awaiting optical character recognition. A further category relates to the proportion of materials in primitive or now defunct digital formats, which are being converted to higher quality or modern standards. However, as we have already observed (3.28), it is apparent that **making an artificial distinction between digitising materials for preservation purposes and to widen their accessibility serves no statistical purpose.**
- 4.18 The actual or perceived “survey burden” that a complex questionnaire presents to any institution needs to be confined within reasonable bounds. A balance must be struck – between simplicity (without which the survey response rate will be prejudiced), and more specific detail (which is required to add value to the analysis). The pathfinder questionnaire appeared to be neither too long, nor so brief as to be too superficial; providing the ambiguity in the definitions are eliminated and better bridging measures adopted for relating the analogue input to the digital output.

5. FOUNDATION DATABASE SOURCES

- 5.1 In order to estimate the total digitisation activity from incomplete survey results, it is necessary to compile other statistics that represent as complete a picture as possible of the total “universe” of cultural heritage institutions. The simplest database to represent that universe would be a complete list provided by each country of their domains: Audio-visual / Film Institutes; Archives / Records Offices; Museums / Galleries; and Libraries. These domains are coarse labels for the institutions concerned, as many could qualify for listing under several headings, and the organisational responsibility is often split differently at ministerial level in each country. Hence a complete list is not yet available. Therefore, we have necessarily resorted to compiling a statistical estimate of the “universe”. We have called this a “foundation database”, because it forms the platform upon which assumptions and sample survey results can be applied to infer estimates of digitisation activity for all the relevant cultural heritage institutions.
- 5.2 The foundation statistics will be made available to Eurostat and the ministries / central statistical offices in each country to verify and update on acceptance of this report. The database also covers countries outside the European Union (EU27), and will be shared with UNESCO as well as those other countries agreeing to participate in the framework for generating digitisation statistics. The database can be interrogated on the numeric web-site and an example of the sort of estimates provided is shown below.

Extract only - Numeric Foundation Database	United Kingdom		Estimated EU total
	On database	Estimated	
General Statistics - Source UNESCO / EuroStat			
Total population 2005	60,218,000	60,218,000	491,400,000
GDP per capita (PPP) US\$ 2004	31,950	31,950	24,938
% of population making at least an annual visit to a:			
Library	49	49	32
Museum / Gallery	42	42	28
% of households with access to the internet	63	63	47
Year of data	2006	2006	2006
Cultural Totals - Source The Economy of Culture in Europe			
Cultural employment 2004 '000s	..	604	4,646
Direct public expenditure on "Culture" € Millions	10,508	10,508	51,558
Year of data	2003	2003	2004
CULTURAL DOMAINS RELEVANT TO THE NUMERIC STUDY			
Archives ● Libraries ● Museums ● Film / Audio-Visual Institutes			
<i>(excludes school libraries, and part of public & special libraries)</i>			
Total Revenue Expenditure of these Domains in € Millions	..	2,444	12,319
Total Staff FTE	39,683	39,683	220,664
Total Number of "Relevant" Institutions	4,748	4,748	45,930

5.3 The database derives from a variety of sources listed below:

- * United Nations Educational, Scientific and Cultural Organisation (UNESCO);
- * International Monetary Fund (IMF);
- * EUROSTAT;
- * ERICARTS⁶;
- * European Group on Museum Statistics (EGMUS) surveys;
- * Association des Cinémathèques Européennes (ACE);
- * Conference of European National Librarians (CENL) survey;
- * LIBECON (EU study to collate statistics on libraries);
- * Initial results from the Numeric survey.

The data sources are attributed in the database on the web-site.

5.4 Many of these data sources were incomplete and, where there were missing values, totals were inferred ("grossed") using averages based on values for available countries, weighting the estimate according to population size and/or GDP. The methodology is described below (at 5.6). It is important to note that these inferred values are intended to facilitate the estimation of EU totals and do not purport to be reliable bases for any single country.

5.5 As more data are assembled, we expect to either refine these estimates or preferably overwrite them with actual data.

Grossing Methodology

5.6 For the purpose of grossing countries within the European Union, we defined two groups, based on their GDP at Purchasing Power Parity in 2004 (sourced from the IMF).

GDP\$ / population (@PPP) 2004

Group 1:	Luxembourg	71,405	Group 2:	Slovenia	20,863
	Ireland	38,986		Czech Republic	19,432
	Austria	32,738		Malta	18,760
	Denmark	32,678		Hungary	16,884
	United Kingdom	31,950		Slovakia	14,666
	Netherlands	31,807		Estonia	14,221
	Belgium	31,557		Lithuania	13,078
	Finland	30,922		Poland	12,893
	Sweden	30,674		Latvia	11,855
	Italy	28,963		Romania	8,417
	Germany	28,413		Bulgaria	8,188
	France	28,252			
	Cyprus	26,609			
	Spain	25,082			
	Greece	22,891			
	Portugal	21,154			

⁶ Council of Europe/ERICarts: "Compendium of Cultural Policies and Trends in Europe", 8th edition 2007. <[http:// www.culturalpolicies.net](http://www.culturalpolicies.net)>

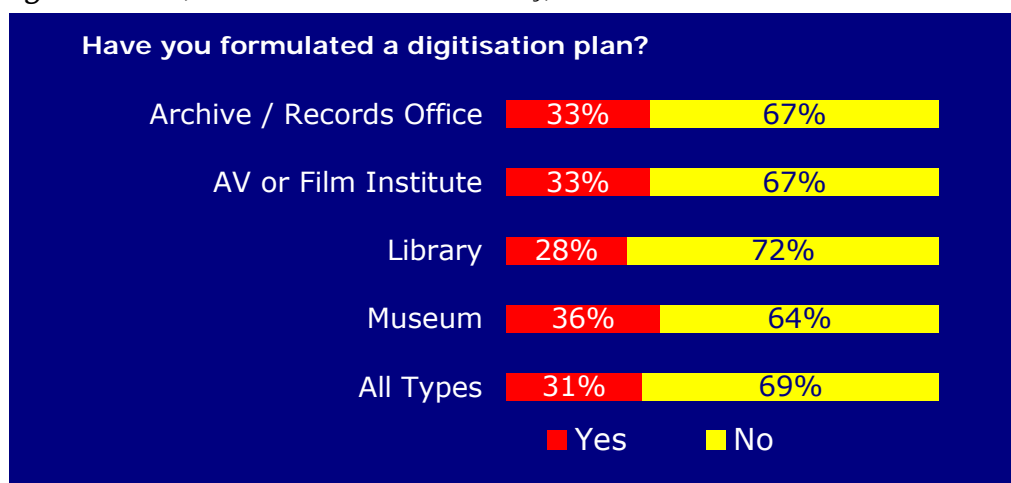
- 5.7 The missing values within the foundation database have been inferred from the known values within each of these groups, based on per capita measures for each heading within the database. Where no data existed for all countries within the group, the EU average was used.
- 5.8 Where possible, other contextual data have been used to inform the estimates: Where we have been able to locate a directory, this has provided a reasonable estimate of the number of relevant institutions, and on this base the number of staff and institutional expenditure has been approximated, applying factors derived from known staffing and budgets in the other countries. An example of this approach is as follows:
- step a* : Country β is in the EU, with GDP (PPP) per head of population greater than €21,000 (as at 2004), and has 4 archives;
 - step b* : The observed average number of staff per archive is 24.4 full-time equivalents (i.e. based on known data);
 - step c* : Therefore, the estimated number of staff for Country β is: $24.4 \times 4 = 97.6$
- 5.9 Financial data are converted to euros using the relevant inter-bank exchange rates before grossing. No attempt has been made to convert to constant prices or to purchasing power parity. Therefore, the estimates represent cash expended per annum.
- 5.10 The current foundation database provides an indicative first estimate of the “population” of cultural heritage institutions, and will be refined as the study progresses. It is accessible on the Numeric web-site, and visitors have been asked to provide feed-back on the estimates. Each cultural ministry and / or national statistical office is asked to scrutinise their data, and provide appropriate statistics to overwrite the estimates.
- 5.11 In the meantime, the foundation database provides a platform upon which to make some overall “sizing” estimates relating to digitisation activities. For instance, the total annual revenue expenditure of the 46,000 relevant institutions⁷ in the EU is estimated at €12.3 billions. By assuming the proportion of total institutional spending devoted to digitisation projects, it is possible to estimate the relevant digitisation costs. Such assumptions can be framed by information about individual digitisation projects accruing to the EDL project and the Numeric study or from survey results. Similarly, observed proportions of collections already digitised can be used to infer the global total.

⁷ Not all “cultural institutions”, as defined by ISO / UNESCO, will hold heritage collections for digitisation – e.g. School Libraries. See sections 7 and 8 for a commentary on defining the “Relevant” number.

6 FRAMEWORK DESIGN

6.1 We believe that to rely on surveys alone to measure the digitisation activity in Europe is not an option. This is why we have invested time in generating a foundation dataset about which to size assumptions when undertaking estimates to stand in the place of survey results. Our experience also leads us to believe that placing undue reliance on a pan-European set of sample surveys would be a mistake. We know that relatively few institutions (less than one-third - see figure 1) have adopted formal project plans to direct their digitisation objectives, and consequently the majority do not maintain the management information in a form that is convenient to report in surveys. This reinforces our belief in a framework informed as much as possible by a survey, but **capable of resorting to other available statistics⁸ to frame realistic estimates if the response to the surveys is insufficient.**

Figure 1 (Source: Numeric Pathfinder Survey)



6.2 We also recognise the difficulty in gaining a consistent response to sample surveys across so many countries. The ministries / central statistical offices in each country have immediate priorities that may not be able to accommodate the implementation of a survey on digitisation activities within the time period we require. Therefore, **we recommend a two tier approach to implementing any sample surveys** about the digitisation activities amongst Europe’s cultural heritage institutions in 2008.

6.3 **The professional associations**, active at the international level, mainly represent nationally prominent institutions, and **form the**

⁸ There are several emerging sources, that can inform assumptions framing estimates: Just a few examples, include: Koninklijke Bibliotheek - Projects 2007-2011, Astrid Verheusen; French digital library project, based within the BnF which is currently scanning the pages of 300,000 books into digital format as part of a plan to set up a European virtual library; The project reports collected by the ADL project and Numeric; BL preservation lifecycle costing projects.

principal tier in the recommended approach. They have been largely successful in gaining survey responses from their members⁹. They will be consulted over the recommended approach to establish whether they are willing to either incorporate the framework questions in their own surveys or are willing to recommend to their members the completion of the Numeric framework survey. Data so gathered from a representative number of such prominent institutions will provide a reliable basis for measuring activity, and these same institutions will be able to advise where other significant effort is taking place in regional and other local institutions.

- 6.4 Gathering information from this second tier of Regional / Local institutions requires the assistance of the proposed "Centre of Excellence", if in situ, and / or the **relevant cultural ministries and / or central statistical offices in each country to coordinate surveys directed at a sample of regional / local institutions**, and they will be canvassed accordingly. As much as possible, the selection of the samples needs to be informed by the "intelligence" provided by the first tier institutions. The deriving survey data, appropriately weighted, can then be combined with results for the national institutions to provide a comprehensive picture across Europe.
- 6.5 The "pathfinder" questionnaire, used to test the approach and collect initial data, is being improved to eliminate ambiguities and clarify definitions as described in the section 7. A lot has been learnt to more consistently measure digital output against the equivalent analogue materials. The framework questionnaire will also be simplified to reduce the burden on responders.
- 6.6 Eurostat will be consulted, and as part of their input, each ministry will be asked to verify the basic statistics relating to the number of institutions in each domain in their country, and to confirm the budgets and associated cultural statistics that the study team have estimated using data collated by ERICarts¹⁰ and surveys conducted by CENL and EGMUS. (This foundation database was described earlier in section 5). This will allow the proper weighting of survey responses, to infer appropriate estimates for missing / non-responding institutions. UNESCO will be asked to recommend the survey to the relevant ministries outside Europe, but at least to those in India, China, Australia, Canada and USA. Regardless, the framework survey questionnaire will be made available on the

⁹ Comprehensive statistical surveys have been established by the Conference of European National Librarians and the European Group of Museums, the results of which have informed the NUMERIC research. The Ligue des Bibliothèques Européennes de Recherche, the Consortium of European Research Libraries and the Association des Cinémathèques Européennes have also assisted NUMERIC. Further assistance will be sought from EBNA (Archives) to cover the professional networks and FIAT (Film Institutes).

¹⁰ ERICarts is the European Institute for Comparative Cultural Research. CENL is the Conference of European National Librarians, and EGMUS is the European Group of Museums.

Numeric website for direct completion, just like the pathfinder survey.

- 6.7 The detailed approach will be refined and international endorsement sought at a workshop to be held in Luxembourg in early 2008. An important outcome from this workshop will be the standard questionnaire, and the commitment of all interested parties, both professional and ministerial to support the framework.
- 6.8 The technical and more detailed aspects of this recommended framework are described in the following sections:
- 7 Review of Definitions / Classifications
 - 8 Defining the "Universe"

7 REVIEW OF CLASSIFICATIONS / DEFINITIONS

Classifications

- 7.1 An important aspect of our work to-date has involved the investigation of appropriate classifications for target institutions to summarise the information when responding to surveys. The classifications we set out to investigate were as follows:
- [i] **objective:** (i.e. the driver for digitisation projects at institutional level)
- * preserve a true representation of the original: mainly audio-visual material;
 - * distribute access to either the most needed part or all the original content.
- 7.2 The consultation and pathfinder survey responses made it clear that these were impractical, if not esoteric considerations and it was impossible to differentiate parts of the collections by a single motive. We have noted earlier in this report (3.28) that a general measure of overall progress will better inform policies, so we have discarded this classification from the framework.
- [ii] **input / output:** (i.e. Grouping of analogue/digital materials)
- * audio-visual;
 - * 3 dimensional objects;
 - * manuscript requiring careful handling / colour management;
 - * normal printed books;
 - * newspapers and other periodicals;
 - * normal one-sheet materials (postcards, photographs, etc.);
 - * difficult flat materials (plans, art objects, maps, etc.).

- 7.3 At the outset, we assumed that it would be pragmatic to provide a detailed breakdown of different types of material in a survey scheme, and that summaries could be made during analysis, by combining headings where there was too much granularity in the returned data. On the whole, the classifications provided in the pathfinder survey proved workable; (see page 2 of the questionnaire available from the web-site). However, it was equally clear that the analogue materials included in a summary group (e.g. Printed matter) might be counted differently to measure the digital output; and, vice versa.
- 7.4 Thus books, in their analogue form, are counted in "volumes", but the digital output is measured in "pages". Most printed matter could be accounted for in equivalent pages, but collections of newspapers are not generally recorded as volumes, and sometimes the analogue statistics relate to "issues", or "editions" of the same "title". Furthermore, the institution is often interested in counting the number of "articles", rather than the number of pages that have been digitised.
- 7.5 Despite this, the only consistent equivalent input/output measure for printed materials is the number of pages, and this standard has been promoted by the CENL in their survey of National Libraries.
- 7.6 Other non-print / images mostly held by libraries, could be a convenient catch-all summary group, but includes very different materials: The analogue measure for Manuscripts (number) is problematic, and the output not always easy to relate to pages. Similarly photographs, engravings, drawings, maps, postcards, and posters are often counted in "Collections" or "Sets", and the number of images may differ in the digitised output files. However, this highlights a problem with the analogue measure rather than the standard for the digital output, and we have concluded that it is important to retain the accepted ISO standards already set for statistical definitions of the materials held in libraries.
- 7.7 The categorisation of materials held by archives, audio-visual institutions and museums is more difficult to reconcile. One respondent to the pathfinder survey commented, "This survey fails, as most surveys, to fit in the reality of a film archive. A national film archive cannot measure 'content' by the hour *or give any other accurate numerical or statistical measures of the collection.*" Another stated that "The survey makes many assumptions about detailed knowledge of and *categorisation of collections which is often impossible in a large and complex organisation of any size and antiquity*". These sentiments serve to show how difficult a uniform classification across all the domains is proving to establish. However, for the framework to succeed in gathering statistics, institutional managers will need to either count or estimate both the

analogue and digital materials in collections, and, apply a universal measure.

- 7.8 In Museums, for instance, the established surveys rarely cover the quantities included in collections – the emphasis is placed on space provided for heritage types: 1. Art, archaeology and history; 2. Science and technology, ethnology; 3. Other. Similarly, the unit of measurement in the Archives domain is the length of the shelves for housing the records. The Audio-visual domain, by contrast, concentrates on very detailed carrier types. We consider the headings defined in the pathfinder survey to have been practicable but recognise the established groupings that EGMUS have based on the UNESCO classification (omitting Zoological and botanical gardens).
- 7.9 For Archives, we have noted the need to translate the metres of records into equivalent pages to derive sensible input / output measures. The classification of the materials between those of historic import and government records may, however, be artificial.
- 7.10 For the Audio-visual domain a unit of time is the single practical standard and the only consistent measure across such a multitude of carrier types. Therefore, despite the apparent difficulty of adopting this measure, we recommend this unit across a classification that includes: Film, Video, and Audio materials, as advocated by PrestoSpace.

[iii] **quality:**

- * still images – (a) text; (b) graphics; (c) mixed;
- * granularity (sound / picture definition);
- * other standards – e.g. with transcription of written text; attached metadata.

- 7.11 As noted in our analysis of the available statistics (6.12), such breakdowns are important in explaining differences in cost. Furthermore, whether purely text or mixed text and image analogue sources are being digitised, **another dimension is required to take account of the quantity of output that has been interpreted by optical character recognition (OCR)**, providing it is accepted that this conversion process falls within the definition of "Digitisation".
- 7.12 There is considerable diversity in both master file and accessible file formats (see appendices V and VI), and any analysis would necessarily need to be restricted to identifying the most frequently adopted formats, but take account of compression standards.

[iv] **heritage type:**

- * museum; archive; library; other;

- * architectural / archaeological / natural / scientific / industrial etc.
- * academic / private / public collection / copyright.

7.13 Six main institutional types were tested in the pathfinder survey to cover the relevant domains: Audio-visual / Film Institute; Archive / Records Office; Library; Ministry of Culture (holding heritage collections); Museum; and "Other type". Most institutions were able to identify with these headings, and within these domain headings it is also possible to distinguish sub-groups (e.g. academic / research libraries).

7.14 The collections of many institutions often cover various types of material, whether natural, scientific, industrial, etc.. But, simplicity is needed upon which to base appropriate sample frames.

7.15 Summarising the above, we propose the adoption of the following classifications for survey and analysis purposes:

Type of Institution

- * Archive / Records Office
- * Audio-visual or Film Institute
- * Library
- * Museum
- * Other inc. Gov. Dept. / Agency

Institutions would indicate the primary heading under which they fall and specify whether they occupy either a National or Regional / Sub-Regional locus.

[v] **usage / access and type of end-user:**

- * limited to physical storage device;
- * closed / limited access, e.g. at the institution;
- * internet;
- * scientific / academic research;
- * educational;
- * commercial;
- * professional / administrative;
- * general open access;
- * local history / community memory.

7.16 **The importance of identifying the use made and / or access to digitised outputs was underlined by the majority of those people we consulted** (see 4.3). The pathfinder survey tested the practicability of gaining these data; and, we have concluded that it is possible to obtain an overall measure of the volume of users / visitors gaining access to the materials. However, given the fact that only 40% of institutions could provide such estimates, we also

believe that it would be difficult for institutions to accurately disaggregate this figure according to the number of internal, paying and open access enquiries. Never-the-less, it is possible to log the access policies they apply and infer the estimated usage accordingly.

Definitions

- 7.17 The UNESCO definitions as developed by ISO provide standard guidance at a generic or headline level to many of the possible survey items. There are some anomalies: For example, "Research Libraries" are not classified separately by ISO; and for practical purposes, the European Group embraces a narrower definition of "Museums", than UNESCO. The latter includes all entities recognised by ICOM, including aquaria, monuments, sites, nature reserves, and zoological and botanical gardens.

Definition of Relevant Population

- 7.18 For the purposes of defining the "Universe" within which estimates of digitisation activity can be made and surveys may be appropriately directed, we propose the following definitions:
- 7.19 **Audio-visual / Film Institutes:** Institutions traditionally responsible for preserving audio-visual collections including corporates, broadcasters, research institutions, and other such organisations. This is restricted to all members of the International Federation of Television Archives (FIAT) and the Association des Cinémathèques Européennes (ACE), as well as those institutions capable of benefiting from the support of PrestoSpace. Archives / records offices, libraries, and museums, would be otherwise classified unless audio-visual materials formed the principal part of their collections.
- 7.20 **Archives / Records Offices:** Institutions whose primary purpose is the preservation of documents of historical and/or rare value, providing these comprise cultural heritage as recognised by the Ministry of Culture in their own country. (For instance, institutions holding land-registries and population census records would usually be excluded) Members of the European Board of National Archivists (EBNA), and archives serving regional or municipal governments would be included.
- 7.21 **Libraries:** These are the institutions defined under ISO:2789 including: National libraries (allows for more than one in a country), Libraries of Institutions of Higher Education, and some Public (i.e. Major / Regional) and some Special Libraries (Government, Professional and Learned Institutions, and Libraries within voluntary organisations, museums, etc.), but excluding: Neighbourhood Public Libraries, School Libraries, and some Special Libraries (Health / Medical Services, Media, Industrial and Commercial enterprises).

(The omission of the latter sectors of libraries is recommended because their digitisation activities either fall outside the scope of cultural heritage, or in the case of small Public Libraries the majority simply provide access to materials. The simple rule of thumb, admittedly arbitrary, that we have adopted at this stage is to include 25% of Public and Special Libraries as relevant institutions in the foundation database. The actual proportion can be varied for each country as the study proceeds and more data becomes available upon which to base the estimate.) "Research Libraries" are included under the heading of "National", "Higher Education" or "Special" Libraries of Regional importance.

- 7.22 **Museums:** These are institutions as defined by ICOM, to acquire, conserve, research, communicate and exhibit material evidence of people and their environment. These are classified by their predominant types of exhibits and collections. For the purposes of the framework, the following are included: Museums of Art, Archaeology and history, Natural History and Natural Science, Science and Technology, Ethnography and Anthropology, a Specialised single theme, Regional and General interest. Institutions excluded are: Aquaria, Monuments, Sites, Nature reserves, and Zoological and Botanical gardens.

Definitions of Materials

- 7.23 Different surveys have variously counted "objects", "exhibits", "artefacts", "volumes", "issues", "images", "records", "discs", "tapes", "reels", "collections", and many other "items". Subject to the matters reported at [ii] above (7.3 : 7.10), the pathfinder measures have proven practicable, but confirmation is required from CENL, LIBER, CERL, EGMUS, EBNA, FIAT, ACE and PrestoSpace to reconcile all needs:-

"Printed type":

Count of *"bound volumes"* and *"pages"*.

"Manuscripts", "Autographs", "Sheet music":

Count of *"collections"* and *"pages"*.

"Images materials":

Count of *"items"* or *"sets"* and number of *"images"*.

"Archives / Records"

Count of *linear "metres"* and *"pages"*.





"Museum Collections":

Count of *"objects"*.

"Audio-visual materials"

Count of *"hours"*.

7.24 The **suggested taxonomy includes the items listed in bold**, in the following scheme, and this also shows how the surveys of each sponsoring organisation cover the materials:

Pathfinder	CENL		
Books (inc "Rare books")	Books	Key to main domains: Libraries  Archives  Museums  Film / Audio-visual 	
Newspapers	Newspapers		
Journals and other serials	Serials		
Government publications			
Other typed print	Other printed material		
Manuscripts	Manuscripts		
Sheet music			
	Autographs		
	Other non-printed matter		
Maps	Maps		
Photographs	Photographs		
Engravings	Engravings		
Drawings	Drawings		
Posters	Posters		
Postcards	Postcards		
Other images	Other image material		
Records of government		Archival material	
Records of historic import			
All other archived records			
Man-made artefacts		Three-dimensional objects Fabrics and Textiles	
Natural world specimens			
Works of Art – 2D			
Works of Art – 3D			
		Art, archaeology history	EG MUS
		Science, technology, ethnology	
Other museum objects		Other	Presto Space
Music / recorded sound	Sound Recordings	Audio	
Film and video recordings	Film and Video Recordings	Film Video	
Any other items	Other		

Definitions of Expenditure

7.25 It is clear that a consistent approach to accounting for the cost of digitising materials would be beneficial, even though in many institutions the ability to report information in surveys to the recommended specificity may be difficult¹¹:

- * The relevant expenditure on digitisation includes the value of all resources directly devoted to preparing materials for digitisation, undertaking the digitisation, and any further work involved in making the content accessible.
- * The cost of any equipment (and media to store the digitised output), made in the year should be reported separately, and only the depreciation charge should be included in the relevant cost of digitisation.

¹¹ Many institutions, responding to the pathfinder survey, were unable to report costs of digitisation because their financial information system did not provide for this level of detail. Therefore, the guideline set out above is provided to facilitate local benchmarking exercises and reviews of processes within the institution, rather than infer that surveys are the only way in which estimates of expenditure can be derived.

- * The relevant cost includes the apportioned cost of in-house staff time, and any expenditure incurred on the use of temporary staff and contractors. The time of IT support staff should be excluded, unless this time relates directly to digitising materials or “loading” content to servers. The estimated cost of staff time, should include the tax overheads borne by the employer, and all remuneration and expenses costs relating to the employee.
- * Where another institution bears the cost and does not make a recharge, the estimated value of the “cost-in-kind” needs to be recognised separately, as the actual cost is plainly not “nil”.
- * Where the expenditure is shared between partners, the appropriate institutional share of the cost should be pro-rated on the volume of digital product.

7.26 As and when we are able to assemble sufficient data to base costing models, these can be made available on the Numeric web-site to assist individual institutions to estimate their expenses according to this guidance. The more detailed guidance already provided on the Numeric web-site will be updated to incorporate the revised definitions.

8 DEFINITION OF THE UNIVERSE

8.1 Based on the UNESCO / ISO definitions, we estimate that there are approximately 227,000 public sector cultural heritage institutions in the EU27 member countries. However, many of these are school libraries (111,000) and, as outlined previously (7.21), a large proportion of Public and Special libraries will be limited to lending services rather than engaging in digitisation activities. We have also observed that many of the archives, libraries and museums hold audio-visual collections, and we have assumed that the number of individual cinema and broadcasting archives is limited to those occupying positions of national prominence. There are also about 2,400 privately owned institutions museums. Therefore, for the purposes of estimating the size of the study universe, there are approximately 46,000 “Relevant” institutions in the EU27:

Domain:	<i>defined at:</i>	Relevant Institutions
Archive	(7.20)	766
Museum	(7.22)	14,299
Library	(7.21)	30,836
AV / Film Institutes	(7.19)	29
All Relevant Institutions		45,930

8.2 Gaining agreement on the definition and size of this universe is important, because this establishes the foundation for all the statistical estimates. i.e. the values within the foundation database reference to this definition of the universe, and these scale the

estimates of digitisation accordingly. We have proposed the following methods for gathering information upon which to inform the scaling assumptions:

- * analysing information that has been and is continuing to be collected on specific digitisation projects, and using these and other sources to frame assumptions for making informed estimates of activity;
- * cooperating with and assisting the international professional associations in undertaking surveys amongst their members, or undertaking the surveys on their behalf;
- * promoting sample surveys beyond the membership of the international professional associations in each country, in order to balance the impression that would otherwise be dominated by nationally prominent institutions.

- 8.3 Given the breadth of the survey exercise and the different manner in which the cultural institutions in each country are organised, it is essential to maintain as simple an approach to the sampling strategy as possible. The sample surveys are, after all, intended to provide reference and context for the overall estimates, and the reliability of these estimates can be better judged on the response actually achieved. Representative responses from the main international associations (CENL, CERL, LIBER, EBNA, ACE, FIAT) in themselves would provide a very solid base, but to add the required context for the wider universe of institutions, the canvass would need to be extended.
- 8.4 A practical way of extending the canvass would rely on the principal tier institutions (those included at 8.3) referring the survey to those exemplars they know at regional / local level. We could then weight the results, to ensure that the survey estimates were adjusted to properly relate to the universe mapped in the foundation database.
- 8.5 We do, of course, recognise the value that the assistance of the relevant cultural ministries and / or central statistical offices could bring to bear on coordinating the sample surveys in their own countries. This would provide for better inter-country comparison and for this purpose we would advocate including responses from about 30 representative institutions for the country as a whole. We have considered a stratified domain / country quota scheme, but based on our experience to-date do not consider that the required level of survey participation could be achieved. Instead, we believe that the more pragmatic approach, outlined above, is better designed to balance the effort required by responding institutions.
- 8.6 The estimated distribution of the relevant institutions for each country / domain, is available from the foundation database on the web-site. Each Ministry / Central Statistical Office will be asked to

review the available directories of the domains for which they are responsible as soon as possible to verify the estimates of the number of institutions we have made.

Desk Research Bibliography – Studies Reviewed

Appendix I

Key:

Thematic group	describes thematic focus of the report (there are 3 groups - see below)
1	the major focus of the report are digitisation activities - this is explicitly formulated in the goals of the survey
2	the major focus of the report is broader than mere digitisation activities
3	the major focus of the report is narrower than digitisation activities and covers only some digitisation-related issues
Target population	points country of origin of institutions that have been surveyed identifies what kind of institutions were surveyed
Scope/significance	identifies geographical coverage and significance of the survey (international, national, local)
<i>international</i>	covers institutions from diverse countries, is commissioned/performed by the international organisations/consortia
<i>national</i>	covers diverse national institutions, is commissioned by national authorities and/or major FALM professional bodies
<i>local</i>	covers a small sample of institutions (in one country), is developed for local needs of smaller professional bodies

Study / Report	thematic (group)	countries	target population	scope & significance
CENL/EDLproject digitisation survey, 2007.	1	AL, AR, AT, AZ, BE, BA, BG, HR, CY, CZ, DK, EE, FI, FR, GE, DE, GR, HU, IS, IE, IT, LV, LI, LT, LU, MT, MD, NL, NO, PL, PT, MK, RO, RU, SM, SP, SK, SL, ES, SE, CH, TR, UA, UK, VA	national libraries	international
Haan, Jos de; Mast Renée; Varekamp, Marleen; Janssen, Susanne. Visit our site: a digitisation of culture [online summary]. 05/04/2007. Available http://www.scp.nl/english/publications/summaries/9789037702965.html , in Dutch http://scp.nl/publicaties/boeken/9037702902/Bezoek_onze_site.pdf	1	NL	cultural heritage institutions, libraries, public broadcasters	national

Study / Report	thematic (group)	countries	target population	scope & significance
Zuraidah Abd Manaf. The state of digitisation initiatives by cultural institutions in Malaysia: an exploratory survey. Library review, 2007, vol. 56, no.1, pp. 45-60.	1	MY	cultural institutions (is not defined, but likely incl. libraries, museums)	national
2006 digitisation survey final report [online]. Metropolitan New York Library Council (METRO). 2006 October. Available http://www.metro.org/images/stories/pdfs/2006_digsurveyreport.pdf	1	US	academic libraries, public libraries, special libraries, school libraries	local
Digital Material in European National Archives [online]. European Board of National Archives. 2006 November. Available http://www.narc.fi/EBNA/docs/EBNA-digireport.pdf	1	AT, BE, CZ, DK, UK, EE, FI, FR, DE, GR, HU, IE, IT, LV, LT, LU, MT, NL, NO, PL, PT, SK, ES, SE	national archives	international
DLF-Aquifer Services Institutional Survey Report: Aquifer Services Working Group Report on the Institutional User-Services Survey Results [online]. 9 March 2006. Available http://www.diglib.org/aquifer/SWGisrfinal.pdf	2	US	libraries (mostly academic)	local
Kahl, Chad M.; Williams, Sarah C. Accessing digital libraries: a study of ARL members' digital projects. In The journal of academic librarianship [online], 2006, vol. 32, issue 4, p. 364-369. Available ScienceDirect.	3	US, CA	academic libraries	international
Manžuch, Zinaida; Pernaravičiūtė, Jurgita. Lithuania: digitisation activities. 2006	1	LT	libraries, museums, archives	national

Study / Report	thematic (group)	countries	target population	scope & significance
Mugridge, Rebecca L. Managing digitisation activities. 2006 September. Available [summary only] http://www.arl.org/bm~doc/spec294web.pdf	1	US	academic libraries	national
Statistische Gesamterhebung an den Museen der Bundesrepublik Deutschland [online]. Berlin 2006. Available http://museum.zib.de/ifm/mat60.pdf	2	DE	museums	national
Status of Technology and Digitisation in the Nation's Museums and Libraries [online]. IMLS. January 2006. Available http://www.imls.gov/resources/TechDig05/Technology%2BDigitisation.pdf	2	US	museums, libraries, archives, state library administrative agencies	national
Bültmann, Barbara; Hardy, Rachel; Muir, Adrienne; Wictor, Clara. Digitised Content in the UK Research Library and Archives Sector [online]. November 2005. Available http://www.jisc.ac.uk/uploaded_documents/JISC-Digi-in-UK-FULL-v1-final.pdf	1	UK	libraries, museums, archives	national
Digitalisierung und Erhalt von Digitalisaten in deutschen Museen. Dirk Witthaut. Unter Mitarbeit von Andrea Zierer, Arno Dettmers, Stefan Rohde-Enslin. - Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, 2005. IV, 116 S. : graph. Darst. http://www.langzeitarchivierung.de/downloads/mat/nestor_mat_02.pdf (Digitisation and the Preservation of Digital Data in German Museums / Dirk Witthaut. In co-operation with Andrea Zierer, Arno Dettmers, Stefan Rohde-Enslin. - Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, 2005. IV, 116 p. : ill: http://www.langzeitarchivierung.de/downloads/mat/02_summary.pdf)	2	DE	museums	national

Study / Report	thematic (group)	countries	target population	scope & significance
Rosen, Mitchell; Frey, Franziska. RIT Survey of American Museums on Digital Imaging for Direct Capture of Artwork [online]. April 2005. Available http://www.cis.rit.edu/museumSurvey/documents/surveyAnalysisReport.pdf	3	US	archives, camera manufacturers, libraries, museums, cultural institutions, private photo studios	local
TAPE survey factsheet [online]. 2005. Available http://www.tape-online.net/docs/Tape_survey_factsheet.pdf	1	PL, DE, FI, IT, ES, UK, NL, RU, AT, SE, HU, DK, IE, LT, SK, NO, CH, BE, CZ, EE, SP, RO, SL, LV, IS, CY, HR, GR, MT, MK, AL, PT, TR	archives, libraries, museums, research institutes, broadcasters, companies, private collectors	international
Preservation and digitisation plans: PrestoSpace questionnaire [online]. March 2004. Available http://www.prestospace.org/questionnaire/index.php?sid=3	2	Europe (further not specified)	audiovisual archives	international
AMICO member museum technology survey [online]. 2003. Available http://members.amico.org/docs/amico.techsrvsrslts.pdf	3	US	museums	local
ASERL survey of member digitisation activities/collections [online]. November 2003. Available http://www.aserl.org/projects/digitisation/ASERL_Digitisation_Survey-Digitisation_Mgrs.pdf	1	US	academic libraries, public libraries	local

Study / Report	thematic (group)	countries	target population	scope & significance
<p>Assessment of End-User Needs in IMLS-Funded Digitisation Projects [online]. 2003 Available http://www.imls.gov/pdf/userneedsassessment.pdf</p>	3	US	academic libraries, public libraries, museums, archives, special libraries	national
<p>A Study into the Key Needs of Collecting Institutions in the Heritage Sector [online]. Final Report 1 May 2002, Faculty of Arts, Cultural Heritage Centre for Asia & the Pacific, Deakin University . Available http://sector.amol.org.au/___data/page/183/keyneeds.doc</p>	2	AU	libraries, archives, museums and galleries	national
<p>A study of business models sustaining the development of digital cultural content [online]. Wall Communications Inc. 4 June 2002. Available http://www.canadianheritage.gc.ca/progs/pcce-ccop/reana/model-study.pdf</p>	2	CA	archives, libraries, research institutes	national
<p>An Open and Interoperable Platform for the Observation and Evaluation (Benchmarking) of Digitisation Projects. Case Study: 1st Benchmarking Practice in Greece [online]. D. K. Tsolis, E. G. Karatzas, D. A. Koutsomitropoulos, V. C. Dourdounis and T. S. Papatheodorou. Available http://www.hpclab.ceid.upatras.gr/viografika/kotsomit/pubs/cr93_en.pdf; http://www.benchmarking.gr/index.php?module=bench1&op=report;</p>	1	GR	digitisation projects (sectors are not specified)	local

Study / Report	thematic (group)	countries	target population	scope & significance
Astle, Peter J. and Muir, Adrienne. Digitisation and preservation in public libraries and archives. In Journal of librarianship and information science [online], 2002, vol. 34, no. 2, p. 67-79. Available SAGE Publications.	2	UK	public libraries, archives	local
Dorner, Daniel D.; Chawner, Brenda; Searle, Samantha. Creating a digital Babylon? Results of a survey of digitisation activities in New Zealand [online]. May 2002. Available http://www.natlib.govt.nz/downloads/Survey_of_digitisation_in_New_Zealand_2002.pdf	1	NZ	archives, tertiary libraries, museums, public libraries, local/regional authorities, government agencies	national
Tanner, Simon; Deegan, Marylin. Exploring charging models for digital cultural heritage: digital image resource cost efficiency and income generation compared with analog resources [online]. June 2002. Available http://heds.herts.ac.uk/mellon/HEDS_charging_models.pdf	2	UK, Europe (countries are not specified)	academic libraries, public libraries, national libraries, museums and galleries, archives, county councils	international
Williams, Esther B. Digital community services: pacific libraries and archives. Future prospects and responsibilities [online]. A report. September 2002. Available http://portal.unesco.org/ci/en/files/6607/10395389010williams.pdf	2	CK, FM, FJ, KI, MH, NR, NU, PG, PW, WS, SB, TK, TO, TV, VU	libraries, archives	international

Study / Report	thematic (group)	countries	target population	scope & significance
<p>Information technology in Canadian Museums: a survey by Canadian Heritage Information Network [online] 2001. Available http://www.chin.gc.ca/English/Pdf/Reference_Library/Information_Technology/it_survey_report.pdf</p>	2	CA	museums	national
<p>EMII European standards survey [online]. 2000. Available http://www.emii.org/map/ques02.htm#1_5 (questionnaire form); source data http://www.emii.org/map/source.htm; analysis http://www.emii.org/map/dk.htm</p>	2	DK, FR, DE, IS, IT, NL, NO, SE, PT, UK	museums	international
<p>Klijn, Edwin; Lusenet, Yola de. In the picture: preservation and digitisation of European photographic materials [online]. Amsterdam, 2000. Available http://www.knaw.nl/ecpa/PUBL/pdf/885.pdf</p>	2	UK, DE, ES, IT, NO, PT, NL, DK, SE, CH, AT, EE, FI, AL, HR, FR, GR, IS, LV, LI, SK, RO, AD, BE, CZ, GE, IE, PL, SL	archives, libraries, museums	international
<p>Garnett, Tom; Gwinn, Nancy E. Preservation and digitisation – natural partners? [online]. 23 August 2001. Available http://www.ifla.org/VII/s19/conf/consws01.html</p>	1	PT, FR, UK, US, FI, LB, NL, JP, KR, NZ, NO, CA, RU, IE, IT	national libraries, academic libraries, public libraries, consortia	international
<p>IFLA/UNESCO survey on digitisation and preservation [online]. Compiled by Richard Ebdon and Sara Gould under the direction of Marie-Thérèse Varlamoff. 1999. Available http://portal.unesco.org/ci/en/files/2172/11715367541Survey.doc/Survey.doc</p>	1	AU, AT, CL, CZ, GR, GL, HK, IS, ID, IR, IL, JP, KR, KE, LV, LT, MY, MT, MN, NL, NZ, NO, PH, MK, SP, SK, ZA, SE, UK, US, YU, ZW	national libraries, public libraries, archives	international

Study / Report	thematic (group)	countries	target population	scope & significance
<p>Towards a learning nation: the digital contribution [online]. Recommendations Proposed by the Federal Task Force on Digitisation. 31 December 1997. Available http://www.collectionscanada.ca/8/3/r3-407-e.html</p>	1	CA	federal government agencies (not specified in detail)	national

ACKNOWLEDGEMENT

We are grateful to the following for their help and for acting as coordinators for the Numeric Study in their own countries:-

Irene Hyna in Austria; Dr Jean Moulin in Belgium; Jakob Heide Petersen in Denmark; Indrek Eensaar in Estonia; Minna Karvonen in Finland; Stefanos Kollias and George Bolanis in Greece; Thorsteinn Hallgrímsson in Iceland; Guy Frank in Luxembourg; Drs. Robert van der Zwan in The Netherlands; Lars Gaustad in Norway; Maria Inês Cordeiro in Portugal; David Dawson in the UK.

We are also most grateful to those persons too numerous to name for their patience and attention in guiding our research.
