

NUMERIC



European Commission
Information Society and Media

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

Study deliverable N° 8:

Study Report

Study findings and proposals for sustaining the framework

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

Chartered Institute of Public Finance and Accountancy (CIPFA)
No 1 Croydon
CR0 0XT
United Kingdom

IPF

May 2009

<http://www.NUMERIC.ws>

Contents of the Study Report

Section		Page
0	Executive summary	1
–	Introduction	6
1	Background to the study	7
2	Approach	13
3	Survey results, context and interpretation	32
4	Review of the methodology	75
5	Proposals for sustaining the framework	85
6	Recommendations and implications	96
	<i>Key word index</i>	99
Appendices	<i>See page 105 for a detailed index to the appendices</i>	
I	Report of survey – technical notes/tabulations	106
II	Database map/programme listings	185
III	Checklist of survey instruments/guidance	186
IV	Classification of survey respondents' comments	214
V	Acknowledgements	229

Figures and tables in the report

Figure:	Page:
1 Sample for the European Union Area (EU27)	22
2 Relative distribution of survey responses	35
3 Institutions possessing a specific budget for digitisation activity	44
4 Formulation of a written digitisation plan/strategy	47
5 Resourcing of budgets and formulation of plans	48
6 Sources of digitisation funding	50
7 Undertakers of the digitisation completed to date	50
8 Summarised progress by type of cultural institution	52
9 Comparative file formats employed for text and images	55
10 Comparative file formats employed for audio-visual materials	56
11 Distribution of costs per page for digitising text and images	58
12 Free access policies compared by type of institution	60
13 Possession of online catalogues	62
14 Proportion of digitised materials available on the internet	63
15 Estimated annual digitisation spend per head of population	70

Table:	Page:
1 Classification of materials held in collections	17
2 Survey management responsibilities in each country	21
3 Institutions per country in the guideline and adopted samples	30
4 Estimated relevant institutions by cultural domain	31
5 Survey questionnaire response rates	34
6 Example weighting calculation	36
7 Questionnaire completion rates: extract 1, for questions 4–16	40
8 extract 2, for questions 17–32	41
9 extract 3, for question 33	42
10 Financial resources identified in budgets	45
11 Full-time equivalent staffing resources	46
12 Digitisation plans	47
13 Sources of funding for digitisation activities	49
14 Undertakers of digitisation projects	51
15 Progress made towards the digitisation of collections	52
16 Progress by the national prominence of the institution	53
17 Progress made towards the digitisation of materials	54
18 File formats adopted for digitising images and text	57
19 File formats adopted for sound and motion images	58
20 Median reported digitisation costs	59
21 Institutional access policies to their digitised materials	60
22 Online catalogue reference to digitised materials	61
23 Digitised materials publicly available on the internet	62
24 Off-line user requests to access digitised materials	64
25 Reported expenditure planned on digitising materials	65
26 Estimated staff resources of relevant institutions	69
29 Annual requests to access digitised materials	71

- Pagination commences overleaf

0 EXECUTIVE SUMMARY

BACKGROUND

- 0.1 Many different types of institution make up the cultural sector, and within this diverse universe, the “Numeric” study has attempted to establish a framework where the progress towards digitising the cultural heritage of Europe can be measured in statistical terms. This involved development and testing of the framework and the associated methodology for the collection and analysis of data on digitisation of material by libraries, archives and museums to better identify the total European digitisation effort and progress. Quantified indicators for measuring the following variables were sought:
- * digitisation costs, investments and funding sources;
 - * volume and growth of digitised resources, related to the analogue collections held by institutions;
 - * the characteristics of digitised outputs, including their formats and user access.
- 0.2 The first year of the study (2007/8) was devoted to reviewing the most appropriate approach for collecting such statistics, and the second and last year (2008/9) was spent collecting data and, in the light of this experience, considering how an ongoing statistical series can be better established to monitor progress in future.
- 0.3 The European Commission funded the study to develop a coherent and eventually comparable framework which would allow the Member States to report on and measure the status quo and progress being made in digitisation. The current study depended substantially on the cooperation of individual experts in various ministries, professional associations and directly involved in managing individual cultural institutions. In this connection, the European Commission’s Member States’ Experts Group was instrumental in nominating a national coordinator in each country. The national coordinators either directly administered the survey, or assisted in its implementation. The study has proven the practicability of collecting relevant data, but has also revealed the need for the development of better definitions and the more consistent engagement of national statistical systems in supporting their adoption.
- 0.4 Statistics provoke controversy simply because they are intended to provide a factual base for the consideration of policies and the evaluation of their success. The current study has generated some

initial data that can illuminate various aspects of the digitisation activities being undertaken across Europe. However, the survey data were generated adopting different sample criteria in different countries, and do not purport to provide a sufficient base for comparisons of international progress. This will be something that can only develop with time, and the improvement in definitions and coverage, that matures according to the lessons learnt with this and future surveys. However, the study has provided some useful benchmarks and insights for many interested in the development of the digitisation of the national cultural heritage. Indeed, the experience has prompted most countries to express a keen desire to collect further data in future. The results of this study provide a platform upon which such future statistics can be gathered.

- 0.5 The survey was designed to be directed to institutions maintaining those collections that would contribute significantly to the value of the nation's digitised cultural heritage. The task of identifying these institutions was devolved to national coordinators according to initial guidance prepared by the study team. In total, this resulted in the identification of 5,752 such European cultural institutions. The survey was then directed at one-quarter of these "relevant" institutions, from which data were returned by 788; reflecting a representative mix (at European level) of archives, museums, libraries, and institutions specialising in audio-visual and other heritage collections. This report concentrates on an analysis of the results for all such institutions summarised for the whole of the European Union area. (All 27 member states participated in the survey, with the exception of Malta). At this level, we can be highly confident that survey results properly reflect the actual position for Europe as a whole¹.
- 0.6 The questionnaire employed on the survey sought detailed information about the institutions' collections and the volume of the materials that had been and were planned to be digitised. Not surprisingly, the response tended to diminish as the level of detail requested increased.

Resourcing of digitisation

- 0.7 Less than half of all survey respondents (48%) declared that their institutions possessed a specific budget with identifiable resources dedicated to digitisation. The value of these budgets summed to €80 millions, representing little more than 1.1% of the overall institutional annual budget. This compares with about 2.5% of institutional staff time being devoted to digitisation work – though these costs have only partially been included in the specific

¹ That is to say there is less than a one-in-ten chance that the range of error about the proportions measured in the survey exceeding $\pm 2.7\%$.

digitisation budgets. Whilst respondents reported that about 30% of the cost of all digitisation so far completed had been funded by grants (provided by central, regional or municipal governments), most work had been funded from the internal resources of the institution – i.e. from their own budget set-asides.

- 0.8 Although the proportion was much higher, particularly amongst national libraries and broadcasters, generally only one-in-three institutions possessed formal written plans for tackling the digitisation task ahead.

Progress towards digitisation

- 0.9 The majority of digitisation work (63%) appears to be undertaken in-house, by the institutions' own staff. A further 6% of the work was completed by lead partners working on behalf of a group of other institutions; notably among archives, film institutes and national libraries. External specialist contractors appear to be handling the significant remaining volume (31%).
- 0.10 The study has formulated a measure of relative "Backlog" of digitisation work still to be undertaken. This highlights the national libraries and relevant university libraries as having more than 20 times the volume of work so far completed, still to be undertaken. This notional measure also reveals significant backlogs for archives, special libraries and broadcasters.
- 0.11 The museums appear to record the highest proportions of their collections so far digitised; and this has led some commentators to suggest that this arises because the respondents have interpreted the recording of materials on digitised catalogues within the definition.
- 0.12 The survey established that the reported rate of progress, even amongst institutions of the same type, differed considerably. For this reason, there is much variation in the statistics; such that the generalised measures, by themselves, would be misleading without taking into account the range of materials held in the collections of cultural institutions.

Digitisation formats and costs

- 0.13 The main master formats currently employed for printed materials and images are TIFF and JPEG files. However, for many this transfer to such formats represents only part of an ongoing digitisation process with for instance, 26% of such files relating to printed text reported as still requiring OCR processing. A much more diverse range of formats for audio-visual processing was reported,

reflecting the different technologies, including upgrade paths for transferring across carrier types.

- 0.14 Against this background there was a very wide distribution in digitisation costs associated with each type of material. The unit costs (per page / per item / per hour) reflected both the different range of processes included as well as the standards of digitisation adopted in each institutional type. For example, 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. Regardless, the data collected provides a useful base, where no previous information existed.

Accessibility of digitised materials

- 0.15 About one-half of all institutions and three-quarters of all libraries operate a free access policy to the materials they have digitised, whether or not these are made available on the internet. Whether or not the “Rights” exist to make such materials available is the significant factor. The majority of libraries and archives possess on-line catalogues to their collections, and of these nearly two-thirds distinguish their digitised materials.
- 0.16 Taking these factors into consideration, the typical proportion of digitised material that has been made publicly available on the internet by libraries is 70%, and nearly 50% by archives. However, we have found it impossible to gain an accurate picture of the volume of requests made via the internet to access the materials. Instead, we have been able to gain an impression of the access made to such materials by persons making physical visits to the institution. 95% of the volume of such requests was reported by the national libraries and the archives. We estimate that this would translate to nearly 43 million annual requests to access digitised materials made on the premises of all the relevant European cultural institutions.

Recommendations

- 0.17 Whilst our initial guidance provided a step forward, it is clear that further guidance is needed to provide for the more consistent identification of “Relevant” institutions. These are the institutions with collections that represent the most significant proportion of the heritage materials that the nation would either wish to digitise or has seen digitised. A dedicated interest group supported by a specialist coordinator needs to review this task urgently.
- 0.18 It is clear that for many institutions the quantification of their analogue collections remains as problematic as tracing their digital

outputs. Generally the institutional management information systems are not yet sufficiently comprehensive to cover digitisation activities. The incentive to record and provide the relevant information needs to be developed for such institutions, and individual ministries in each country could better recognise this priority.

- 0.19 Streamlining future survey questionnaires to reduce the burden on those institutions providing the data is essential. This will also assist those collecting the data, because there will be a reduction in the number of persons within the institution with whom to correspond regarding the completion the questionnaire.
- 0.20 The alternative exists to incorporate the collection of the required data in national policy or other statutory returns. This is desirable, but it is essential for each country to adopt the common definitions such as those developed for the Numeric study questionnaire, otherwise the uniform base for comparison will be lost. However, some urgency relates to the further development of these definitions (for instance the way in which newspapers are counted and internet requests for information are logged), rather than the process adopted for collecting the information. The task needs focus and could be accommodated by the same task group approach suggested for improving the definition of “relevant” institutions.
- 0.21 Any future survey should be kept simple enough to enable a broad majority of the targeted institutions to collect and provide the requested data with only limited additional effort involved. At the same time, future work will have to accommodate the need in some countries for more detailed benchmarking as well as the needs and interests of the professional associations such as CENL representing the institutions implicated in digitisation of cultural heritage.

INTRODUCTION

Digitisation is allowing ‘memory institutions’ (archives, libraries, museums, broadcasting and film institutes) to diversify and enrich the information economy by providing online access to Europe’s cultural heritage. The Numeric study has sought to establish a framework to measure the pace and cost of translating their collections into this rich resource.

The study was conducted over two years (May 2007 to May 2009), and this report outlines the findings of a survey conducted among cultural institutions in member states of the European Union; and, based upon this experience, sets out recommendations for monitoring their digitisation activities in future. Future surveys can benefit from the lessons learned from the current study to improve the logistics for international cooperation in the definition and collection of relevant data.

A large volume of data has been generated by the surveys conducted during the study. As a result, it has been possible, for the first time, to provide a sound, consistent and statistically founded base for tracing progress within the European Union. At this level, the range of error about the survey estimates is well within the range of confidence accepted for policy and similar market research purposes (see 3.14). Even so, we must introduce a note of caution concerning the interpretation of the results. The statistics describe a universe of cultural organisations that by their very nature often represent unique collections. Therefore, the statistical measures attempt to describe a real cultural sphere, where one of its attractions is its diversity. Added to this, in each country, the size and representation of the sample has differed, and the reader is asked to treat all the summary tables with corresponding circumspection.

As with any new statistical series, further refinement of the definitions can now take place, and attention can be directed towards incorporating them in cultural institutions’ management information systems.

The report concentrates on describing the survey findings and makes high-level summary estimates for the European Union area as a whole. National analyses have been undertaken and shared with each individual country participating in the framework; interested parties may discuss these more detailed statistics with the relevant national coordinators, in the agencies listed within the report (see Table 2).

1 BACKGROUND TO THE STUDY

- 1.2 The Numeric study was established to design and implement a statistical framework to gather information about current and planned digitisation of cultural material for the European Commission². This report describes the approach adopted; the results of a survey conducted among cultural institutions in each member state; and proposals to develop the statistical framework that has been established in future years.
- 1.3 Before the study began, in May 2007, there had been no attempts capable of providing sufficiently consistent and comprehensive statistical information about the progress made in any single country in Europe. While there have been and continue to be annual exchanges of experiences about individual digitisation projects, more than anecdotal evidence was required to gauge the progress actually being made in each country to digitise their cultural collections. In addition, although these exercises may have been essential to the policy and annual reporting cycles, they were not yet capable of producing statistical outputs on a comprehensive or consistent basis.
- 1.4 A significant step was taken in 2007 by the Conference of European National Librarians (CENL), to collect statistics, and this helped to inform part of the design of the approach adopted in the Numeric study, but even this survey was confined to a single type of institution. Another long standing series of surveys in the United States of America³ has collected data about the digitisation activities of libraries and archives, but confining the survey to only these institutions would limit the perspective of the extensive activity being undertaken across the whole of the European cultural sector. Other benchmarking initiatives have been undertaken, notably by the Prestospace project and by Training for Audiovisual Preservation in Europe (TAPE)⁴, but focusing solely on audio-visual collections.
- 1.5 What the Numeric study has uniquely established is a reach across all the cultural domains, covering all the major types of cultural institution, and a framework for coincidental and consistent adoption at the international level. But what is the importance of this statistical framework and why should it be further developed?

² Recommendation 1, on the digitisation and online accessibility of cultural material and digital preservation, of 24 August 2006.

³ *Status of Technology and Digitization in the Nation's Museums and Libraries*, US Institute of Museum and Library Services, January 2006.

⁴ *Tracking the Reel World*, 2008.

- 1.6 Reliable measures (statistics) are essential to better inform the way in which digitisation projects are managed at the institutional level, and the way that policies for supporting digitisation of the national cultural heritage are coordinated. From a policy point of view, the following questions remain relevant simply because those European institutions responding to the study survey reported investment of €80 millions annually in the digitisation of their cultural collections, inferring a significant level of expenditure within the whole of the European cultural arena:
- * How much has been spent, and how much more is needed?
 - * What proportion of digitised collections can be read on the internet?
- 1.7 These questions cannot be answered immediately, but a statistical series built up over time will provide the vital currency to inform the evaluation of individual and collective progress.
- 1.8 In the meantime, this study has sought to establish the tramlines for such a statistical framework. Accordingly, the scope of this framework first had to be defined. The cultural collections held by libraries, museums and archives of all sorts contain diverse materials including the collective memory of print (books, journals, newspapers), photographs, museum objects, archival documents, monuments and audio-visual material, to name but a few. Within this complex universe, the statistical measures are intended to exclude 'born-digital' materials (published directly in electronic formats). In other words, the study looks at the baseline for analogue-to-digital conversion of physical materials.
- 1.9 Given this background, we first undertook a review of the research conducted to date to establish the most appropriate framework design for the study, and we concluded that:
- * 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 – indeed, no baseline had been established by such surveys.
 - * These previous studies attempted to trace the perspective from either the 'top down' (policies that relate to institutions in common) or 'bottom up' (management information needs for the institutions). Both appeared to focus on the need for resources, but both approaches failed to define consistent measures for either financial or technical resources. This remains a difficult challenge considering the different ways in

which institutions are organised nationally, let alone internationally.

- * Most effort was 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, but an established statistical framework would be required to allow such definitions to mature.
- * There were no uniform approaches on how to classify digitised collections. Few studies have been able to successfully 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 has been a significant achievement of the Numeric study, although more time is required to embed these measures in the institutional management information systems.
- * The studies we reviewed revealed different levels of aspiration in digitisation work. The basic level suggested that an institution ignored the social or cultural benefits of digitisation beyond its own organisation; an intermediate level referred to digitisation for online access and to preserve the materials; and the highest level set out to exploit the digitised collections by developing educational, tourism, research and other such benefits. Encompassing such diverse aspirations in an invitation to participate in a statistical framework was recognised as a challenge.
- * The International Standards Organisation (ISO) continues to develop consistent definitions in many areas. However, despite work being undertaken to define improved digitisation activity in libraries (ISO 11620), there were no suitable definitions in other institutions, such as film archives. In any event, we recognised the need to define consistent measures for investment in digitisation activities and recurrent expenditure on servicing electronic access and services.
- * Indicators of user needs, usage of materials, accessibility and usability of digitised content needed to be more visible in digitisation surveys, and so we acknowledged this in our own approach.

1.10 Based on this review of previous approaches, we concluded that the framework should be designed to:

- * recognise the different quality standards and processes adopted by different types of institution, but focus on developing high level classifications capable of summarising the major types of digital output;
- * limit the statistical measures or indicators to data that could be substantially collected at present, instead of burdening institutions with a framework that was aspirational rather than immediately practicable;
- * concentrate on headline measures for the main types of institution, and provide for individual national initiatives to develop benchmarking surveys should more detailed statistical information be required;
- * implement a method for making the task of collating consistent data simpler in each country.

1.11 At this initial design stage, the study was well informed by suggestions provided by more than 100 individual consultees, and the views of acknowledged national experts in the field of digitisation⁵. The more extensive rationale for establishing the current framework was rehearsed with the European Commission at the outset of the study, and then the refined aims were confirmed at a workshop attended by interested experts from each of 26 member states, in April 2008. All present subscribed to the importance of establishing the costs to institutions, and the progress made towards both preserving and making their materials accessible online. Such statistics were required to:

- * better identify the total European digitisation effort and progress;
- * stimulate further digitisation by demonstrating the current progress;
- * better inform stakeholders that have an interest or direct involvement in digitisation policies and funding (governments, statistical agencies, cultural institutions, academic and scientific institutions, publishers, industry).

⁵ As nominated by the Member States' Expert Group on Digitisation and Digital Preservation [MSEG] convened by the European Commission.

1.12 While considerable discussion took place between the invited experts concerning the size and design of a questionnaire to collect the desired data, it was agreed that the key statistics sought should describe:

- * digitisation costs, investments and funding sources;
- * volume and growth of digitised resources, related to the analogue collections held by institutions;
- * the characteristics of digitised outputs, including their formats and user access.

1.13 A desire was also expressed at the time to make international comparisons, between countries, although the study has proven the risk of doing so without first recognising the significant cultural and organisational differences that exist between institutions in different countries, or indeed different administrations. Regardless, the survey results provide a sound base for estimating the overall patterns and extent of digitisation activity in the European Union as a whole. At this level, the range of error about the survey estimates is well within the range of confidence accepted for policy and similar market research purposes (see paragraph 3.14).

1.14 It will also be evident that by their very nature statistics usually beg more questions. They are measures confined within the straight-jacket of precise definition, but before an established historic series is available their plausibility is more difficult to judge. The base position established in this study, and described in this report, needs to be interpreted in this context. A base set of statistics has been created which now raises questions that demand further research. The statistical definitions and survey standards will be more easily refined, founded on the valuable experience of this study which has been implemented by 26 of the 27 Member States of the European Union.

1.15 At the conclusion of the study, the overall findings were reported, at a workshop held in Luxembourg, to those National Coordinators of the surveys undertaken in each country. It was generally agreed that further development of the definitional base was required and that action should be taken to provide for this. The overall approach to the collection of data in future was endorsed, but in the meantime some sensitivity was required to ensure that the results of the current study were more valuable in informing that development, rather than providing a definitive base of the progress being made towards the digitisation of cultural collections. The significant points to emerge were:

- * **Benchmarks:** For a large number of Member States, the survey provided them with very useful benchmark data, very often for the first time in the surveyed areas. Therefore, a continuation of this exercise was strongly supported.
- * **Choice of Sample:** The participants agreed that the choice of the sample and the related definition of "relevant" institutions form a central challenge for the comparability of the national results, as on a national level the relevancy was defined from a national point of view. A European survey needs to find a balance between the need to create samples that allow a comparable statistical analysis between Member States and/or cultural sectors on the one hand and the institutional contexts specific for cultural sectors or the national cultural institutions. It was agreed, that further work was required to create a common ground for the survey, while remaining open for exceptions and differing contexts.
- * **Complexity of the Questionnaire:** Many national coordinators encountered problems during the survey, due to the length and complexity of the questionnaire, which was not easy to complete and often required the involvement of different departments and colleagues to answer all points. As a remedy, the study proposes the use of a shorter questionnaire in future that specifically focuses on the headline data requirements. The study provides a draft version (see 5.18) that would then need further and continuous refinement.
- * **Motivation for participation:** Furthermore, the discussions showed that among many institutions there was a lack of motivation for participating in the survey, as it was not always sufficiently clear to potential respondents, how they might benefit. This issue was identified as a crucial determinant in the development of the statistical framework. A solution to these challenges will most likely depend on the approach taken in the national context. Relating national data needs to existing statistical surveys, involvement of national statistical institutes and better usability of the survey by individual institutions were named among the possibilities.

2 APPROACH

- 2.1 As described above, the study was in effect faced by a blank canvas; at the start of the study, we could find no country that had undertaken a comprehensive survey of the digitisation work undertaken across all cultural domains. The base information did not exist at a higher organisational level than the individual institution. The 'primary' data would need to be collected from individual institutions in each country. To ensure that these data could be collected at broadly the same time, and to consistent definitions, a common framework was required. This framework also needed to focus on collecting the data from the 'correct type of institutions', and from a representative sample of these.
- 2.2 With the assistance of the MSEG, nominees from each country⁶ met in Luxembourg to review the recommended approach. At this meeting those attending agreed to confirm the nomination of appropriate national coordinators in each country to implement the survey to generate the baseline data. Considerable discussion focused on three aspects of the approach:
- * the definitions and questions to adopt as the standard for the survey;
 - * the identification of those institutions whose collections would represent significant value to the nation's digitised cultural heritage;
 - * the size and representation of the sample of institutions to include in the survey.

⁶ All countries except Malta were represented at the workshop, held on 2 April 2008.

2.3 Each of these aspects is reviewed in more detail later in this report. In overview, however, the agreed approach can be summarised as follows.

Step 1 National coordinators would verify the total number of cultural institutions in their countries, and identify those that were 'relevant' in the sense that digitisation of their collections would significantly enhance access to the country's cultural heritage⁷. A specific statistical return was distributed to all concerned to record this information as consistently as possible. The importance of identifying these relevant institutions is explained at 2.15.

Step 2 Guidance notes were distributed so that a representative sample of relevant institutions could be identified to participate in the sample survey, to be conducted in the latter half of 2008. The data so collected would be used to estimate the total digitisation activity for all the relevant institutions in each country.

Step 3 Prior to mounting the survey, national coordinators would agree the survey management responsibilities with the study team, so that the appropriate overall coordination and support could be assured.

Step 4 The survey results would be analysed and used to inform estimates made by the study team at a pan-European level, and to the coordinators in each country to inform any national review of the digitisation efforts, achievements, expenditures and digitisation plans in the wider international context.

2.4 The collection of data required specific survey instruments for coincidental application in all countries: standard guidance on how to select a representative sample of institutions, and a common questionnaire. While the overall guidance was adopted to various degrees (to suit national circumstances), the concurrent implementation of the survey by national coordinators in 26 countries established a remarkable network. In about one-half of all

⁷ These were the main archives, film and broadcasting institutes, libraries and museums. It was accepted that judgement, based on local knowledge, was required to decide which among the country's institutions fell into this category. Institutions administered by regional governments and municipalities may have held significant collections of value to the nation's heritage. However, there were clearly some that were irrelevant (school libraries) and insignificant (neighbourhood or small public libraries). In larger countries the proportion of institutions that were relevant was likely to be less than in smaller countries. Records offices/archives, the main museums and libraries of cities, and all film and broadcasting institutes were likely to be relevant. Not all higher education establishments were relevant, since the 'heritage' materials were more likely to be concentrated in the longer established universities.

countries the coordinator represented the appropriate ministry responsible for cultural digitisation policy.

STANDARD QUESTIONNAIRE

- 2.5 The questionnaire, reproduced in Appendix III, was developed after extensive testing and consultation. A 'pathfinder' (or test) survey was conducted in the autumn of 2007 to assess the questions and the survey approach. This was a useful practical procedure to put the prior desk research into proper context, and this led to several refinements.

Funding and resourcing issues

- 2.6 The formulation of a digitisation plan by an institution appeared to be the exception rather than the norm. Consequently, no common management information frameworks covering the digitisation process existed within most institutions upon which to assemble consistent financial management information about their digitisation work. This implied that however meticulously we were to define financial expenditure on digitisation, few institutions would yet have the experience to implement the measures. Also, more general budgets often did not identify resources set aside specifically for digitisation work, even though this may be how the money under a broader vote was to be spent.
- 2.7 Therefore, we proposed to collect information about 'funding', 'budgeting' and 'spending' under several different questions in the questionnaire, to account for the fact that where one institution could provide financial data in one form (e.g. a unit cost), another would only be able to estimate costs for completing a planned digitisation project. Therefore, despite the apparent attraction of shortening the questionnaire in recognition of the difficulty for an institution to provide financial data, we maintained the basis for arriving at estimated statistics from different angles in the survey design.
- 2.8 Furthermore, some institutions had observed how difficult it was to distinguish between IT resources and the proportion of this financial allocation devoted to digitisation. Others reflected on problems in apportioning the cost of staff time to digitisation activity. For this reason, we modified our approach by setting out the definitions in more detail on the questionnaire, but providing responders with the option to describe any specific elements missing from the financial data they provided. In this way, we believed that a recognised standard definition would eventually develop after several survey iterations, but in the meantime, somewhat coarser survey estimates were better than no statistics at all.

Input–output measures

- 2.9 A difficulty soon emerged in the design of the survey about how to measure digital output in a way that could relate to the analogue materials. The Conference of European National Librarians (CENL) survey, undertaken for the EDL project, first encountered this problem, and suggested the use of a proxy measure (average pages per book). We adopted a similar approach in the Numeric questionnaire, specifically requesting primary data in both forms (pages and volumes). After considering similar difficulties in measuring archived records, we also opted to collect data about the average number of pages held per metre of shelf space, but extended the option for institutions to report their analogue collections in 'metres of shelf space', 'volumes' or 'number' of materials.
- 2.10 Measures of digitised outputs, such as the 'number of images/files/pages', reflect the technological process of conversion, and it was necessary to introduce standards that could be readily applied to large numbers of different types of material held in collections. Our solution to this problem was to set just a few measures that could describe both the analogue input and digital output for most types of material: collections measured according to 'hour' for audio-visual materials, 'pages' for manuscript or other types of document, and 'objects' or 'number' for other physical items. The overall design improvements made to the input–output section, between test survey and actual implementation, are summarised as follows.

Table 1 Classification of materials held in collections

Pathfinder categories	Units	Framework categories	Units
Books (including 'rare books')	<i>Volumes</i>	Archival records	<i>Metres</i> ¹
Newspapers	<i>Volumes</i>	Rare books	<i>Volumes</i>
Journals and other serials	<i>Volumes</i>	Other books	<i>Volumes</i>
Government publications	<i>Volumes</i>	Newspapers	<i>Issues</i>
Other type printed material	<i>Number</i>	Serials	<i>Volumes</i>
Manuscripts	<i>Number</i>	Manuscripts	<i>Number</i>
Maps	<i>Number</i>	Sheet music	<i>Number</i>
Photographs	<i>Number</i>	Microforms/microfilms ²	<i>Number</i>
Engravings	<i>Number</i>	Maps	<i>Number</i>
Drawings	<i>Number</i>	Photographs	<i>Number</i>
Posters	<i>Number</i>	Engravings/prints	<i>Number</i>
Postcards	<i>Number</i>	Drawings	<i>Number</i>
Sheet music	<i>Scores</i>	Posters	<i>Number</i>
Other images not classified above	<i>Number</i>	Postcards	<i>Number</i>
Archived records of government	<i>Metres</i>	Paintings	<i>Number</i>
Archived records of historic import	<i>Metres</i>	Any other two dimensional object	<i>Number</i>
All other archived records	<i>Metres</i>	Three dimensional works of art	<i>Objects</i>
Man-made artefacts in museums	<i>Artefacts</i>	Man-made artefacts	<i>Objects</i>
Natural world specimens	<i>Objects</i>	Natural world specimens	<i>Objects</i>
Works of art – two dimensions	<i>Exhibits</i>	Other objects in collections	<i>Objects</i>
Works of art – three dimensions	<i>Exhibits</i>	Monuments ³	<i>Number</i>
Other objects in museum collections	<i>Objects</i>	Film	<i>Hours</i>
Film and video recordings	<i>Hours</i>	Video recordings	<i>Hours</i>
Music and other recorded sound	<i>Hours</i>	Audio (music & other recorded sound)	<i>Hours</i>
Other items not classified above	<i>Number</i>	Other items not classified above	<i>Number</i>

Notes ¹ The respondent was free to state another measure – 'volumes' or 'number'.
² If not shown included elsewhere on the form.
³ Institutions solely responsible for monuments were asked to complete another form.

End-user considerations

2.11 During our consultations considerable emphasis was placed upon the need to gather information on the average number of users or visitors gaining access to digitised materials according to the access policies adopted by the institution. Several improvements to the test survey design were incorporated, but we can acknowledge now that recognised standard management information headings need to be universally adopted (not just defined) before institutions can more consistently report the use made of their digitised materials. In the meantime, we sought information about:

- * whether online catalogues identified digitised materials;
- * the proportion of digitised materials made available to the public on the internet;
- * the number of user requests for digitised materials via (a) the internet, and (b) any other access method (e.g. CD-Rom).

2.12 An attempt to gain further consistency was made in framing estimates of 'user accesses' per annum, rather than for shorter periods which were subject to greater seasonal difference.

General design points

2.13 Further refinements to the questionnaire were informed by two consultation exercises; the first conducted by correspondence with stakeholders, and the second with national coordinators and experts attending the workshop to launch the framework⁸. A number of points arose from these exercises of which the following influenced the design of the final form of the questionnaire.

- * It was apparent that making an artificial distinction between digitising materials for preservation purposes and to widen their accessibility serves no statistical purpose.
- * 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 demanded more information on the usage made of digitised resources.
- * It was agreed to integrate monuments into the statistical framework, providing this would not undermine the successful implementation of the approach across the core collections described in other parts of the questionnaire.

2.14 The finalised questionnaire adopted a design which attempted to demand high-level information at the beginning, followed by more detailed information about the types of material (both analogue and digital) later. This was intended to present a less intimidating approach to institutions that would undoubtedly need to undertake a significant amount of primary research to complete the survey. A further section, seeking information about individual projects, was retained to provide context to the interpretation of the statistics. This information has also been made available to Europeana.

⁸ The same meeting referred to in footnote 6, held in Luxembourg on 2 April 2008.

STANDARD GUIDANCE

- 2.15 A preparatory task undertaken in each country (before the survey was mounted) involved identifying the relevant institutions in which the digitisation of materials would lead to the most significant progress towards the digitisation of the nation's cultural heritage. This was a task that only the national coordinators in each country were competent to undertake, and indeed, a desk-top exercise undertaken during the early stages of the study proved inadequate in properly identifying the core of institutions that housed significant collections that needed to be digitised.
- 2.16 To overcome these earlier difficulties with the use of estimates deriving from the desk research, we needed to gather the data directly. To facilitate the consistent recording of such information, we designed a high-level statistical survey that was undertaken among national coordinators, as described in the guidance they were given, and reproduced in Appendix III. The questionnaire was an administrative pro-forma, called 'Sheet A', requesting the number of all cultural institutions by type in the country and the corresponding number of 'relevant' institutions. Consequently, the definition of which institutions were the most relevant to building of the nation's digital heritage was specific to each country, and the statistical results reflect this varied base; this is discussed further later in the report (see section 4).
- 2.17 There remains a need to standardise the definition to provide for more consistent comparisons in future years (see section 4). In the meantime, the pro-forma used for identifying relevant institutions was combined as a sample selection tool with accompanying detailed guidance notes to all national coordinators. A verified total number of cultural institutions in each country was thereby derived, which identified those that were relevant, either in the sense that digitisation of their collections would significantly enhance access to the country's cultural heritage, or in that these were the most significant institutions currently engaged in digitising the nation's cultural heritage.
- 2.18 Following this, the national coordinators selected a sample of institutions to include in the survey, in the knowledge that these might be used to estimate the total digitisation activity for the nation. In summary, the study team retained responsibility for EU coordination of the survey including:
- * advice to those acting as coordinators in their country and/or for their domain, as well as those persons implementing the sample surveys, if these are different;

- * web and e-mail distribution of survey instruments;
- * collation, validation and analyses of the surveys and other data.

2.19 National coordinators retained responsibility for:

- * the identification of relevant institutions in their own country;
- * sample selection for the survey;
- * translation services, when required, to better communicate the questionnaire and definitions;
- * providing help with queries as they arose from survey respondents;
- * chasing late survey responses.

2.20 Ministries or central statistical offices retained the option to either:

- * act as the national coordinator, but sanction the study team to manage the survey on their behalf; or
- * entirely manage the national framework.

2.21 Based upon these responsibilities, printable versions of the questionnaire were distributed to all national coordinators early in May 2008, to provide them with at least six weeks' lead time to translate the material before the electronic questionnaires were despatched. The electronic survey questionnaire supplied by the study team was a spreadsheet, with built-in choices for answering category questions, and other cross-tabulation and plausibility checks. This was formulated in English, but in many cases translations were made by national coordinators (see Table 2). A covering letter in the national language explaining the objectives of the survey was supplied by most national coordinators, to emphasise the national importance of the exercise. All the survey instruments were also posted to the study website, including the translations, as and when these were received.

2.22 The guidance note circulated to all national coordinators was distributed at the end of April 2008. The note provided a consistent set of rules for preparing for, and undertaking, the sample surveys; it included technical advice on identifying relevant institutions and drawing representative samples, as well as flowcharts to explain the respective responsibilities for those persons administering the required procedures. A 'frequently asked questions' (FAQ) facility

was added to the website to provide guidance on interpreting and using the survey instruments, and to mitigate the risk of inconsistent interpretation of the instruments and questionnaire as the survey progressed.

2.23 The division of responsibilities for the management of the survey in each country is shown below, where ‘national’ indicates the local collation of questionnaire returns. This also shows the language of the questionnaire circulated in the country, and the national coordinators’ agencies, from which they could present a sense of national authority to the survey.

Table 2 Survey management responsibilities in each country

Country:	National coordinating agency:	Survey manager:	Questionnaire language:
Austria	Federal Ministry for Education, the Arts and Culture	IPF	GBR
Belgium	Scientific & Technical Information Service	National	FRA; NLD
Bulgaria	IST Directorate, State Agency for IT and Communications	IPF	GBR
Cyprus	Higher Technological Institute	National	GBR
Czech Rep	National Library	IPF	CZE
Denmark	Library Agency	IPF	GBR
Estonia	Ministry of Culture	National	EST
Finland	Ministry of Education	IPF	GBR
France	Ministère de la Culture et de la Communication	National	FRA
Germany	Stiftung Preussischer Kulturbesitz	National	DEU
Greece	National Audiovisual Archive	IPF	GRC
Hungary	Ministry of Culture and Education	National	HUN
Ireland	Library Council	National	GBR
Italy	Ministry of Cultural Heritage and Activities	National	GBR
Latvia	Culture Information Systems State Agency	National	LVA
Lithuania	Baltic Audiovisual Archival Council	IPF	LTU
Luxembourg	Ministry of Culture, Higher Education and Research	National	FRA
Malta	Heritage Malta	-	-
Netherlands	Digital Heritage Netherlands	National	NLD
Poland	National Library	National	POL
Portugal	National Library	IPF	PRT
Romania	Centre for Studies & Research for Culture	National	ROU
Slovak Rep	Ministry of Culture	National	GBR
Slovenia	Ministry of Culture	IPF	SVN
Spain	National Library/State Archives/Ministerio de Cultura (Museos)	National	ESP
Sweden	National Archives	IPF	GBR
UK	Museums, Libraries and Archives Council	National	GBR

SAMPLE DESIGN

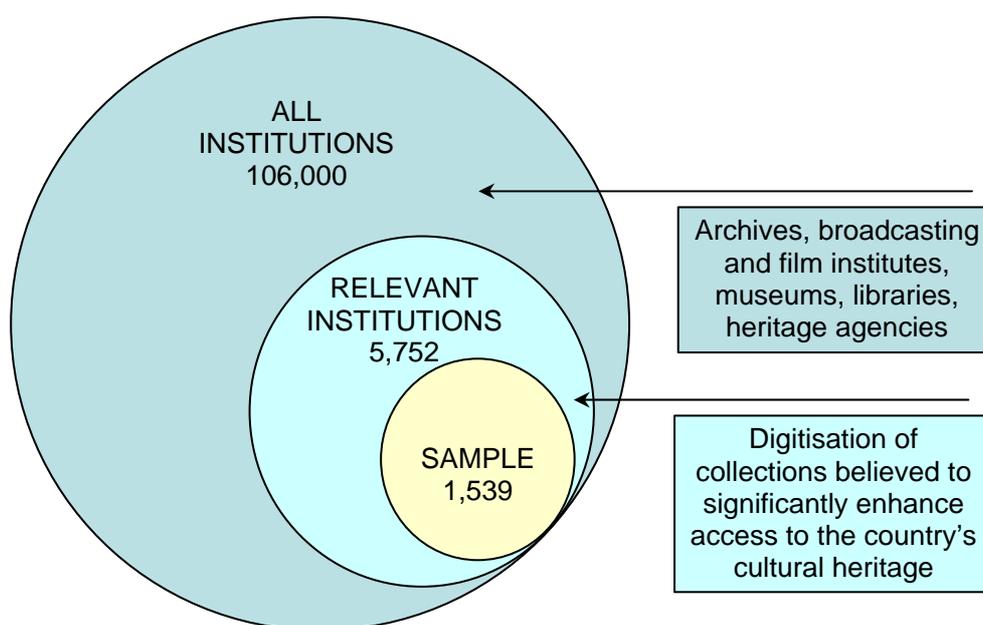
2.24 The sample survey was intended to gain representative data upon which to project an estimate of the digitisation activity in all cultural institutions throughout Europe, and sufficient in individual countries to provide reasonable national summaries. The results of the

'pathfinder' survey in 2007 showed that a low questionnaire response rate was likely unless each institution recognised the national importance of participating. The importance of establishing a national link for the survey was therefore crucial and best accomplished by central endorsements from the appropriate national ministries.

2.25 Furthermore, we anticipated the need for individual institutions to develop their own management information systems to cover digitisation activities, in order to lessen their research burden before responding to the questionnaire. We therefore consulted on an initial proposal for a compact survey approach, based upon a minimum sample of 30 institutions per country, providing there were that number of relevant institutions comprising the significant national digital heritage base. We expected that national coordinators could better assure a full response from this compact sample.

2.26 The national experts consulted at the workshop to launch the framework, held in April 2008, proposed a more ambitious approach, particularly among the larger nations, where there was a desire to measure regional or community differences within national boundaries. Therefore we proposed a guideline sample quota for each country, including sub-quotas for the major types of institution (domains). These guidelines were applied variously by individual national coordinators to suit individual local circumstances (see above). The approach used for the overall sample for the EU27 area is shown in Figure 1.

Figure 1 Sample for the European Union Area (EU27)



- 2.27 The base of relevant institutions is an important part of the statistical methodology, since it has been used to adjust the survey results to reflect the real distribution of institutions in each country, rather than the pattern of survey responses. The base number of relevant institutions was also important in establishing national estimates using the survey statistics.
- 2.28 As the diagram in figure 1 shows, the relationship between the assessed number of relevant institutions and the sample selected to survey would be consistent in each country if the same approach was adopted towards their identification. Specific guidance was issued in an attempt to gain a uniform approach but, at this first stage in establishing the framework, the expediency of using established national policy criteria in some countries meant that different approaches were adopted. In some countries the identification of relevant institutions and the selection of the sample concentrated on those known to have commenced digitisation work.
- 2.29 In others the institutions identified included those with potential collections worthy of digitisation, although the extent of the work was unknown or had not yet started. For this reason, inter country comparisons of the results would be particularly unreliable, although at a summary level comparisons between institutional types are more robust, because the survey results reflect the weightings for all such relevant institutions, regardless of the national samples. However, it is worth reflecting that some distortion may arise because of the mixed methodology applied in each country. The extent of the different approaches adopted in each country is described below.

COORDINATORS' INTERPRETATION OF THE 'RELEVANT' UNIVERSE AND SAMPLING

- 2.30 The national coordinators have reported their own approaches to the selection of the samples and their estimates of 'relevant' institutions, and these observations have a bearing on the interpretation of the survey results.
- 2.31 AUSTRIA: The sample selection was focused on federal and regional institutions and some smaller local institutions with important collections. The absence of a survey response from a major federal museum which undertakes a comprehensive amount of digitised material needs to be noted. The museum consists of 11 big collections each with a separate director.

Those institutions undertaking only limited digitisation and possessing no digitisation plan did not respond to the survey, even though they were thought to possess important collections. Small institutions that did not respond may hold no concise documentation about their collections.

2.32 BELGIUM: The questionnaire was distributed to 283 institutions⁹ in order to provide a reasonable base upon which to provide results for the different communities within the country (nine federal institutions; 39 in the French speaking community; and 235 in the Flemish community). However, no questionnaires were distributed to the German speaking community.

The criteria for selecting the samples were different for each community. The Flemish institutions were selected according to their value to the community's cultural heritage in general, and/or whether they received government funding, in which case they were deemed to be 'relevant'. If the institution was known by the Observatory of Cultural Politics to be actively engaged in digitisation, then it was deemed to be 'relevant' in the Walloon region.

The federal institutions were included because of their size and significance. Given this spread, no undue bias is considered to be inherent in the survey results.

2.33 BULGARIA: The 'relevant' institutions were considered to be those serving a significant proportion of the population, and therefore omitted those with small collections. Within this target, a representative sample was selected by type and region. In meeting the quota, it was necessary to include a number of regional institutions of lesser significance, but this ensured that the sample was truly representative.

Although the pattern of response was even, it was noticeable that a number of non-respondents did not undertake any significant digitisation activity.

2.34 CYPRUS: Some uncertainty attaches to the criteria adopted for the selection of the 'relevant' and the sample institutions. However, the response to the survey was close to the guideline quota of ten sample institutions.

⁹ The initial sample, and the one adopted for computing the rate of response, included nine federal institutions, 30 within the French community, and 190 in the Flemish community (229 in total).

- 2.35 CZECH REPUBLIC: The guidelines were implemented to identify those institutions housing collections of significant value to the nation's cultural heritage. The sample was drawn according to the guidelines, and provided for a representative selection of all types of national and regional institutions.
- 2.36 DENMARK: A tight definition was adopted, identifying 13 of the estimated 520 institutions in the country as 'relevant'. These were all state-owned and the focus was placed on the major institutions. The sample then included all of these. Therefore, the survey results, while providing a complete response, may omit further digitisation activity that is taking place in the wider universe of institutions in the country.
- 2.37 ESTONIA: The 'relevant' institutions were considered to be those housing collections of significant value to the national cultural heritage, and the response to the survey was complete to a representative sample.
- 2.38 FINLAND: The 'relevant' institutions were identified by the Ministry of Education to include those which held collections of national significance to the cultural heritage of the country. A selection covering national and regional representation was made to compile the sample.
- 2.39 FRANCE: Institutions identified as 'relevant' by the Ministère de la Culture et de la Communication were selected according to information returned in the annual reports of institutions, and the French MICHAEL database, 'Patrimoine numérique', which identifies digitised collections and other sources. For the purposes of defining the wider universe of 'relevant' institutions (as shown in Table 3 below), the study team derived the data from the 'Chiffres Clés 2008', published by the Ministry.

The selection was confined to those already undertaking digitisation projects, and the sample was confined to 30 institutions to allow sufficient scope to gain a reasonable response. Ten institutions of each of the following domains were selected: archives; museums; libraries. These were the 'major' institutions in terms of their demand for resources, and staff dedicated to digitisation activities. It will be noted, however, that this excludes any broadcasting or major audio-visual institutions. The Ministry considers that the omission of a wider selection of institutions from the sample does not effect the representation of the survey, and that the results will

not have been unduly distorted by those that did not respond to the survey.

- 2.40 GERMANY: Nearly 800 institutions were identified as being potentially 'relevant'. The survey was directed to a representative selection of 105 institutions (including 66 libraries).
- 2.41 GREECE: The 'relevant' institutions were identified according to those that had contributed digital material to MICHAEL and/or those that had received funding to digitise materials. While this was considered a practicable approach, providing a reasonable universe in which to measure activity, one in five of the sampled institutions responded to the survey.
- 2.42 HUNGARY: The guidelines were implemented to identify those institutions housing collections of significant value to the nation's cultural heritage. The sample was drawn according to the guidelines, and provided for a representative selection of all types of national and regional institutions.
- 2.43 IRELAND: Institutions which hold collections which would contribute to the digitised heritage of the country, but which may or may not necessarily have digital collections at this point, were considered to be 'relevant'. Within this universe a random and representative sample was selected. However, the response omitted three libraries, all of which were prominent national institutions, and two of which have been engaged in more digitisation activity than many of the other Irish institutions.
- 2.44 ITALY: The 'relevant' institutions were considered to be those falling under the administrative responsibility of the Ministry of Culture. These institutions include 424 museums (including responsibility for monuments and archaeological sites), 47 libraries and 100 archives. The sample was limited to 30. The target for the sample, and the response, therefore, omits the potentially wider universe of institutions not directly administered by the Ministry.
- 2.45 LATVIA: 'Relevant' institutions were deemed to be those housing collections of significant value to the nation's cultural heritage. The sample was drawn according to the guideline quotas, and provided for a representative selection of all types of national and regional institutions.

- 2.46 LITHUANIA: All types of institution have been defined as 'relevant', and the sample has been targeted to include a regional spread as well as those most active at a national level.

One of the non-responders was a prominent broadcaster that has been conducting a large scale digitisation project on TV material. Its omission will impact on the results for audio-visual media digitisation efforts and initiatives.

- 2.47 LUXEMBOURG: The principal institutions in the Grande Duchy were identified according to the guidelines, and the Ministry of Culture, Higher Education and Research aimed to collect data separately using existing statistical survey returns. Subsequently the Numeric survey questionnaire was circulated.

- 2.48 MALTA: While Heritage Malta was unable to participate in the current survey, the agency has expressed a preference to implement the survey in future years. In the meantime, the number of relevant institutions in the database has been estimated by the study team.

- 2.49 NETHERLANDS: The 'relevant' institutions were those known to be actively engaged in digitisation, rather than those that may possess collections of national significance. The list was compiled from the pooled knowledge of the Digitaal Erfgoed Nederland (DEN), and the national knowledge centre on ICT and cultural heritage in the Netherlands. The data were gathered using a locally designed questionnaire and a contractor was appointed to assemble the data accordingly.

An initial quota of 30 institutions was drawn, from which 20 responses were received, and a second survey was directed to provide coverage of more than 160 institutions. The national coordinator is satisfied that the survey results give a representative picture for the Netherlands, however, it should be noted that several of the survey questions were either omitted or asked in a different form, so some of the base data in the Numeric scheme are either omitted or have been estimated, according to the discretion of the DEN.

- 2.50 POLAND: 'Relevant' institutions were deemed to be those housing collections of significant value to the nation's cultural heritage. The sample effectively included most of these, and was drawn according

to the guideline quotas, and provided for a representative selection of all types of national and regional institutions.

2.51 PORTUGAL: The main criteria for identifying 'relevant' institutions were the possession of cultural heritage collections, the existence of unique or rare items in the collection or the known digitisation experience of the institution. The sample was then directed at the recognised major institutions and those with digitised collections.

There were a number of significant institutions that did not respond to the survey, including major institutions of the following types: a radio and broadcasting institute; a film archive; three prominent libraries; and four regional archives. However, the overall sample was considered sufficient.

2.52 ROMANIA: The sample was chosen as a representative quota from the identified 'relevant' institutions. These were deemed to be those housing collections of significant value to the nation's cultural heritage.

2.53 SLOVAKIA: The 'relevant' institutions were deemed to be those housing collections of significant value to the nation's cultural heritage. The sample, however, was confined to the nation's principal institutions.

2.54 SLOVENIA: 'Relevant' institutions were considered to be those where digitisation of collections significantly enhances access to the nation's cultural heritage. These included the national institutions but also regional and municipal institutions.

The sample adopted the stratified quotas for the different types of institution. Where few were required within the quota, there was an inevitable focus on major institutions, whereas the more numerous types of institution provided for the inclusion of all administrative responsibilities (national, local, municipal) and a wide regional representation.

The majority of non-responders were regional and municipal institutions.

2.55 SPAIN: The National Library identified 101 relevant libraries, and targeted all of these in the survey. The Ministry identified 104 relevant museums, and circulated the questionnaires to 22 nationally prominent institutions and asked agency managers of the

museums in the autonomous communities to administer the survey for the remainder. However, the identified sample omitted archives because the Ministry did not include these institutions in their survey, and therefore, the study team have estimated the number of relevant institutions, and EU summary results have been used to estimate this domain in Spain. (See 3.77 where the method of grossing the survey results is described). Inevitably the overall survey results will include estimates for archives, where the institutional base will need to be verified by the Ministry, but in the meantime, the estimates reflect a definition of this base that is comprehensive of the cultural sector.

- 2.56 SWEDEN: It was considered important to include a wide spread of small, medium, large, public (state and communal) and non-public institutions in the 'relevant' list. From these, the sample focused on institutions that were considered capable of responding to the survey. The possibility of bias arising from non-response has been discounted.
- 2.57 UNITED KINGDOM: Invitations to participate in the survey were promoted to institutions through the Minerva and associated networks, but without any formal introduction. The study team subsequently identified the 'relevant' institutions adopting criteria mainly based upon the accreditation of cultural institutions, and limiting public libraries to those listed as main branches, serving large populations and open for more than 60 hours per week. Based on this identification a revised sample (103) was identified including a higher level of national institutions, and a regionally representative number of other 'relevant' institutions. The response was reasonably representative, but omitted the considerable activity undertaken in a major library.
- 2.58 In the context of the above points made by the national coordinators, the selection of sample quotas within this universe of identified relevant institutions is summarised in Table 3. It will be evident from the table that the number of 'relevant' institutions per head of population diminishes as the population size of the country increases. However, the guideline samples were fixed according to the absolute number of relevant institutions, and for the larger countries we advocated a larger sample to ensure that it was sufficient and representative.

Table 3 Institutions per country in the guideline and adopted samples

	Total population in millions (Eurostat) 1/1/2008	Relevant institutions	Relevant institutions per million population	Relevant as % of all cultural institutions	Guideline sample	Adopted sample	Adopted sample as % of relevant institutions
Austria	8.332	71	9	4%	49	49	69%
Belgium	10.667	120	11	5%	93	229	191%
Bulgaria	7.640	37	5	2%	35	35	95%
Cyprus	0.789	108	137	96%	10	4	4%
Czech Republic	10.381	237	23	4%	102	102	43%
Denmark	5.476	14	3	3%	13	13	93%
Estonia	1.341	29	22	3%	11	11	38%
Finland	5.300	95	18	7%	29	29	31%
France	63.753	477	7	6%	104	30	6%
Germany	82.218	801	10	5%	105	105	13%
Greece	11.214	48	4	3%	43	43	90%
Hungary	10.045	439	44	22%	47	47	11%
Ireland	4.401	86	20	31%	26	26	30%
Italy	59.619	578	10	5%	41	30	5%
Lithuania	3.366	15	4	1%	14	14	93%
Luxembourg	0.484	12	25	8%	9	9	75%
Latvia	2.271	142	63	13%	24	24	17%
Malta	0.410	5	12	8%
Netherlands	16.405	166	10	9%	102	131	79%
Poland	38.116	119	3	2%	100	100	84%
Portugal	10.618	94	9	13%	83	83	88%
Romania	21.529	565	26	4%	105	105	19%
Slovakia	5.401	129	24	5%	62	62	48%
Slovenia	2.026	180	89	47%	16	16	9%
Spain	45.283	265	6	4%	108	127	48%
Sweden	9.183	212	23	24%	55	55	26%
United Kingdom	61.186	708	12	6%	103	60	8%
EU27	497.455	5,752	12	5%	1,489	1,539	27%

2.59 The reported relative distribution of 'relevant' institutions by type of organisation varied considerably between countries. Even so, museums were the most numerous type of relevant institution in 21 countries and, together with libraries, made up more than four-fifths of all relevant institutions within the EU27 area. This needs to be kept in mind when considering the survey results (see 3.83), given that the overall summary figures will be heavily influenced by this predominant type of institution in the overall make-up of the cultural sector.

2.60 Table 4 summarises the identified 'relevant' institutions by main cultural domain. Although the individual characteristics of the survey samples applying in each country were often different (i.e. some were confined to the major institutions), the overall survey results for the aggregated area of the European Union are scaled to the identified universe of 'relevant' institutions. That is to say, that the characteristics reported by respondents to the survey have been attributed to the complete number of institutions in proportion to their national distribution. Therefore, at the summary level, the

survey approach has provided a reasonable reflection of pan-European digitisation activities, against which individual countries may compare their own survey results in the light of both the local institutions included, and the response rates they achieved.

Table 4 Estimated relevant institutions by cultural domain

	Archives and records offices	A-V and film inst's	Museums	Libraries	Other insts †	All domains
Austria	15	7	33	14	2	71
Belgium	24	9	54	15	18	120
Bulgaria	3	4	15	14	1	37
Cyprus	1	6	37	60	4	108
Czech Republic	29	1	162	41	4	237
Denmark	1	3	8	2	-	14
Estonia	1	2	21	5	-	29
Finland	17	2	60	15	1	95
France	154	27	101	156	39	477
Germany	17	3	528	249	4	801
Greece	11	8	15	6	8	48
Hungary	85	4	184	165	1	439
Ireland	3	3	31	49	-	86
Italy	100	3	424	48	3	578
Lithuania	2	1	6	6	-	15
Luxembourg	2	1	5	1	3	12
Latvia	2	3	110	27	-	142
Malta	1	-	2	2	-	5
Netherlands	50	5	73	32	6	166
Poland	8	3	18	88	2	119
Portugal	18	2	50	22	2	94
Romania	20	1	230	312	2	565
Slovakia	47	2	63	14	3	129
Slovenia	7	2	102	68	1	180
Spain ‡	55	2	104	101	3	265
Sweden	56	4	91	60	1	212
United Kingdom	119	1	227	360	1	708
EU27	848	109	2,754	1,932	109	5,752

† In most countries these 'other types' comprised specialist organisations, where the principal activity did not fall conveniently under any of the main institutional headings. In France, the 'Services du patrimoine' (archéologie et inventaire) accounted for the number shown. In Belgium, these other types tended to be heritage centres, theatres and technical or administrative organisations.

‡ In Spain, the identified sample omitted archives, and therefore, the study team have estimated the number of such relevant institutions.

3 SURVEY RESULTS, CONTEXT AND INTERPRETATION

3.1 This section covers the following topics.

Context	<i>Paragraph:</i>
Rate of response	3.3
Adjusting for the pattern of response	3.7
Quality of response	3.14
Survey results	
Budgets devoted to digitisation, and the proportion of the total budget of the institution this represents	3.31
Full-time equivalent staff time devoted to digitisation work	3.36
Proportion of institutions that have adopted formal digitisation plans	3.40
Sources of funding for digitisation undertaken to date	3.43
Who has undertaken the digitisation work, expressed as a proportion of all work completed to date	3.44
Progress towards digitisation of institutional collections	3.45
Most common formats for the master files relating to digitisation work undertaken	3.52
Unit costs (per page/hour) of digitised materials	3.56
Access policies and the percentage of digitised materials publicly available on the internet	3.60
Cost of planned digitisation work according to type of material	3.69
Universal estimates	
Projection method	3.72
Projected estimates	3.81
Inputs	3.82
Outcomes	3.85
Analysis of survey comments	
General comments	3.89
Digitisation partners	3.96
Digitisation contractors	3.98

[Please refer to the index for references to specific topics – e.g. “audio-visual” materials.]

CONTEXT

3.2 The majority of survey questionnaires were circulated during July and August in 2008, although some were earlier (Slovakia in May) or later (Germany in December). The survey included two types of questionnaire – a ‘generic’ form covering all types of institution, and a supplementary pro-forma for use by agencies with a responsibility for monuments. Two of the 14 returns received in respect of monuments were added to the database; the others either supplemented information provided on the generic questionnaire, or had been completed by institutions not responsible for monuments. Furthermore, we have included the questionnaire returns provided by institutions in response to the ‘pathfinder’ survey, after eliminating those that also completed the current generic questionnaire.

Rate of response

3.3 The following response to the survey was received, from which it will be seen that, for the larger countries in general, the deriving data represents a more limited proportion of their universe of ‘relevant’ institutions.

3.4 The level of response can be (and needs to be) measured in different ways: 51% of the adopted sample; 14% of the relevant institutions. For smaller countries, a higher level of response than for larger countries was preferable, simply because the estimation error based on small samples tends to be higher. This aspect is explored in more detail under *Quality of response* below, since it should be kept in mind that providing the data collected are from a reasonably representative number of the full population of relevant institutions, a lower response rate simply diminishes the level of confidence we can attribute to the survey estimates. In appendix I, we have reported the standard errors attaching to the survey results at national, and sectoral levels, and provided an exemplification on judging the sufficiency of the survey statistics in each. We have also previously referred to the relationship between the number of relevant institutions identified in each country, and the sample size adopted (see 2.28); it follows that a more sufficient response in the larger countries would have been preferred, given the more substantial resources at their disposal.

3.5 Thus, the pattern of response also needs to cover a reasonably representative distribution of the types of institution, to eliminate bias in the survey results, and regardless of the size of the country, the survey results have to be regarded in the context of the population of relevant institutions. Allowing for the variable response between different countries, the overall pattern and

volume of questionnaire returns provides for a reasonably representative and sufficient base upon which to infer estimates for the whole of the European Union (see Figure 2). Less certainty attaches to such estimates for individual domains (main types of institution), but the base is sufficient to provide for reasonably robust comparisons at the EU level.

3.6 The overall pattern of response at the EU27 level reflects reasonably well on the identified distribution of 'relevant' institutions, as shown in the following diagram. ('Film institutes' also include other audio-visual archives and broadcasting institutes.)

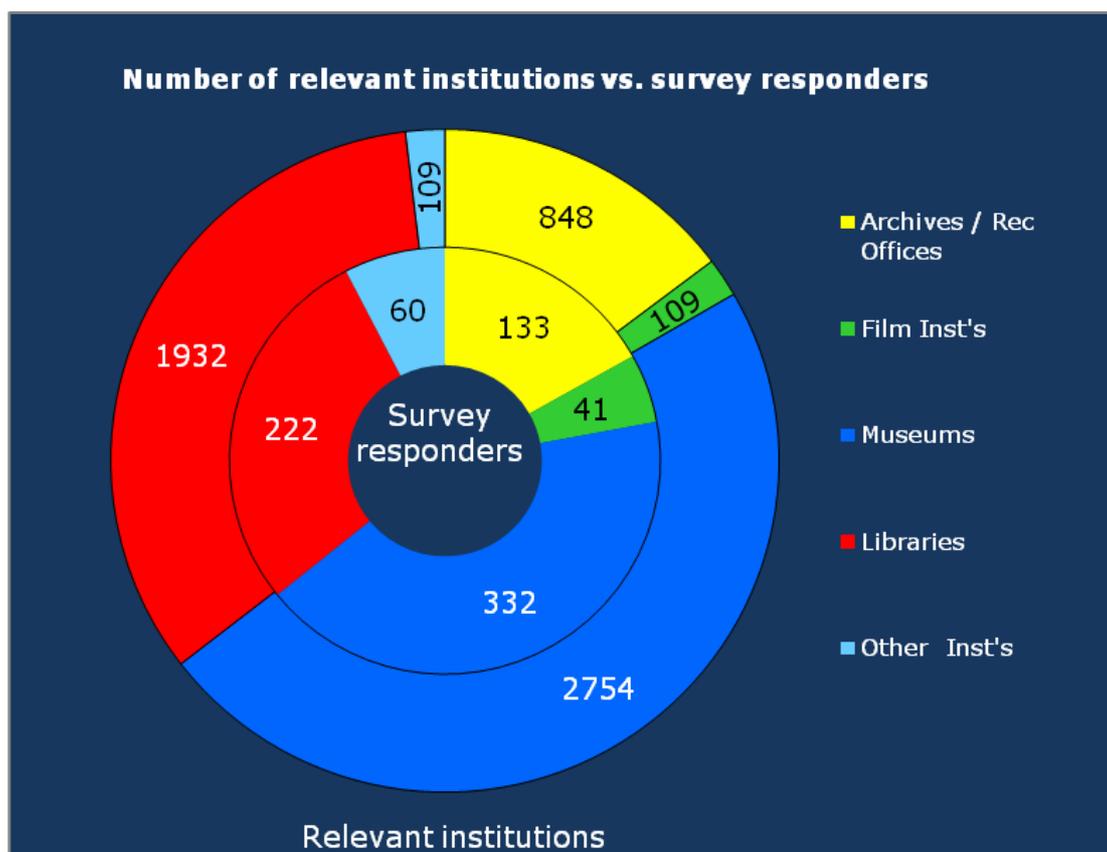
Table 5 Survey questionnaire response rates

	Relevant institutions [1]	Adopted sample [2]	Sample of rel'vnt insts [3] =[2 / 1]	Survey responders [4]	Response rate [5] =[4 / 2]	[6] =[4 / 1]
Austria	71	49	69%	26	53%	37%
Belgium	120	229	191%	92	40%	77%
Bulgaria	37	35	95%	12	34%	32%
Cyprus †	108	4	4%	7
Czech Republic	237	102	43%	28	27%	12%
Denmark	14	13	93%	13	100%	93%
Estonia	29	11	38%	12	109%	41%
Finland	95	29	31%	19	66%	20%
France	477	30	6%	28	93%	6%
Germany	801	105	13%	35	33%	4%
Greece	48	43	90%	21	49%	44%
Hungary	439	47	11%	52	111%	12%
Ireland	86	26	30%	16	62%	19%
Italy	578	30	5%	22	73%	4%
Lithuania	15	14	93%	8	57%	53%
Luxembourg	12	9	75%	2	22%	17%
Latvia	142	24	17%	10	42%	7%
Malta ‡	5
Netherlands	166	131	79%	131	100%	79%
Poland	119	100	84%	32	32%	27%
Portugal	94	83	88%	26	31%	28%
Romania	565	105	19%	39	37%	7%
Slovakia	129	62	48%	11	18%	9%
Slovenia	180	16	9%	10	63%	6%
Spain	265	127	48%	86	68%	32%
Sweden	212	55	26%	26	47%	12%
United Kingdom	708	60	8%	24	40%	3%
Archives	848	262	31%	133	51%	16%
A-V/broadcasters	109	60	55%	41	68%	38%
Museums	2,754	457	17%	332	73%	12%
Libraries	1,932	690	36%	222	32%	11%
Other types	109	70	64%	60	86%	55%
All domains	5,752	1,539	27%	788	51%	14%

† No verified sample identified

‡ Malta did not participate in the survey arrangements.

Figure 2 Relative distribution of survey responses



Adjusting for the pattern of response

3.7 When using the survey statistics to project estimates for all institutions, we have weighted the questionnaire returns to account for the proper distribution of types of institution by the known number that are relevant in each domain, and country. This process ensures that the survey results properly reflect the composition of relevant institutions, regardless of the pattern of questionnaire responses from each type of institution. It also diminishes the unintended bias resulting from either poorly selected sample institutions or a patchy response from particular types of institution. However, it will be observed that like any other logical approach, this adjustment is reliant on accurate information about the number of relevant institutions in the first place.

3.8 Given the relatively representative pattern of response for the aggregated surveys, any adjustment to the results at the EU27 summary level leads to reasonably small changes in the deriving estimates. However, in order to summarise the results at national

level or any other smaller geographic area, the weighting of the survey summary was necessary.

3.9 This is best illustrated in the example calculation set out in Table 6, which examines the digitisation progress made by institutions, as a proportion of their total collections. The table is set out in two parts:

- * the first represents the simple analysis of the survey results (i.e. the universe of institutions responding to this question);
- * the second is the adjustment of the survey summary for all types of institution (i.e. to reflect the real universe of institutions).

Table 6 Example weighting calculation

Part (i)					
Type of institution:	Responders to question	No need to digitise	Already digitised	Outstanding digitisation	Total collection
	N ^o	%	%	%	[2]+[3]+[4]
	[1]	[2]	[3]	[4]	[5]
Archive/records office	98	41.6%	9.4%	49.0%	100.0%
Audio-visual or film institute	20	36.9%	9.5%	53.6%	100.0%
Broadcasting institute	9	15.6%	10.8%	73.7%	100.0%
Museum of art, archæo, hist	190	12.4%	30.3%	57.3%	100.0%
Museum of science, tech, ethn	21	19.0%	33.3%	47.7%	100.0%
Other type of museum	71	15.0%	27.8%	57.2%	100.0%
National library	26	45.6%	3.0%	51.4%	100.0%
Higher education library	42	64.7%	6.1%	29.2%	100.0%
Public library	54	56.8%	10.8%	32.4%	100.0%
Special or other type of library	50	43.9%	9.2%	46.9%	100.0%
Other type of organisation	35	19.9%	24.4%	55.7%	100.0%
Total responding institutions	616	30.2%	19.3%	50.5%	100.0%

Part (ii)							
Type of institution:	Relevant total	No need to digitise	Already digitised	Outstanding digitisation	No need to digitise	Already digitised	Outstanding digitisation
	N ^o	[a]x[2]	[a]x[3]	[a]x[4]	[b]/[a]	[c]/[a]	[d]/[a]
	[a]	[b]	[c]	[d]	[e]	[f]	[g]
Archive/records office	848	353	79	416	41.6%	9.4%	49.0%
Audio-visual or film institute	99	37	9	53	36.9%	9.5%	53.6%
Broadcasting institute	10	2	1	7	15.6%	10.8%	73.7%
Museum of art, archæo, hist	1,074	133	326	616	12.4%	30.3%	57.3%
Museum of science, tech, ethn	1,020	194	339	486	19.0%	33.3%	47.7%
Other type of museum	660	99	184	378	15.0%	27.8%	57.2%
National library	64	29	2	33	45.6%	3.0%	51.4%
Higher education library	773	500	47	226	64.7%	6.1%	29.2%
Public library	778	442	84	252	56.8%	10.8%	32.4%
Special or other type of library	317	139	29	149	43.9%	9.2%	46.9%
Other type of organisation	109	22	27	61	19.9%	24.4%	55.7%
Total relevant institutions	5,752	1,948	1,128	2,676	33.9%	19.6%	46.5%

3.10 In the example set out in Table 6, valid survey responses to the question about digitisation progress have been received from 616 institutions. These responses by type of institution are not in the same proportion as the actual mix of the total number of relevant institutions: 5,752 in column [a]. Since 'relevant' institutions represent the real universe within which the sample is intended to

portray digitisation activity, their relative distribution (sic) can be used to place the appropriate 'emphasis' on the survey results to more appropriately reflect that reality. In other words, the results for each type of institution are weighted to adjust the results to reflect the actual position, rather than the pattern of response.

- 3.11 So, where the average part of the collections that do not need to be digitised for all survey responders was 30.2%, in column [2], the weighted proportion reflecting the true average in column [e] would be the sum of the product of the survey responses in column [2] and the weights, in column [a]. The sum of these products provides a more representative proportion of collections deemed not to require digitisation (33.9%), since it is based on the distribution of relevant institutions in the EU reported by all national coordinators. This process is also applied to the remaining survey responses (in columns [3] and [4]) to derive the weighted estimates.
- 3.12 The same method applies to weighting the national results for all types of institution. It will be observed that at the EU27 level, the weighting makes little difference, and this provides some comfort in the actual representation of the survey response by type of institution.
- 3.13 It is also evident from the weighting methodology that the base of relevant institutions is a critical factor in establishing reliable estimates of the overall digitisation activity. See Table 3, where smaller countries generally have more relevant institutions per head of population, and a higher proportion of all their institutions that are considered to be significant in terms of the digitisation of the national cultural heritage. This simply reflects the relative scale of each country so that, regardless of its size, the reliability of the survey results depends upon the survey's representation of the nation's relevant institutions

Quality of response

- 3.14 As indicated above, we can be confident that the summarised survey results represent a reasonably true picture for the EU27 area. The survey errors at this level compare well with the

standards adopted for market research studies that are published to inform policy reviews or report on public opinion¹⁰.

- 3.15 'Confidence' is a relative term but, when applied to statistics, it relates to the possibility that the actual position is outside the range suggested by the survey results. Therefore, if we wish to increase our confidence in the results, we need to accept the possibility that the actual range of error about our estimate increases. For instance, the range of error, when based on a 90% level of confidence, is $\pm 2.7\%$, at most¹¹, but it increases to $\pm 3.2\%$ at a 95% confidence level. Containing the range of error within these limits provides us with a level of confidence in the results that is comparable with the standard for research studies.
- 3.16 The range of error about the estimates for all domains is, of course, narrower than it is for individual types of institution, and it follows that the survey error generally increases as the number of institutions in the category diminishes and the sample (as a proportion of the relevant institutions) reduces. An exemplification of the possible errors about the survey estimates for individual domains and types of institution is set out in the report of survey. Regardless, a sense of proportion is required about the interpretation of statistical errors, particularly if the actual number of relevant institutions is small, and the sample response is representative.
- 3.17 However, given the initial base established by the survey, further information collected at a later date will consolidate the statistical information at the national level. Furthermore, there are a variety of organisational differences that exist between countries, and, at this stage in the development of the framework, separate analyses of the respective national results compared with the summary findings for the EU27 area have been provided directly to national coordinators (i.e. not attached to this report).

¹⁰ The overall response to the survey exceeds the number of returns received in all other previous studies relating to digitisation activities. Furthermore, the number of cases provides a sufficient base to limit the error about the estimates with a similar level of confidence that most market research and other academic studies regard as acceptable. See paragraph 3.15 – which refers to the range of error in this regard. In cases where it is important to distinguish fine differences, such as voting intentions, an error of $\pm 3\%$ is accepted by most polling organisations. Audit studies are often based upon results where the error may be $\pm 5\%$.

¹¹ 'At most', because the error about an estimate of a proportion diminishes as the percentage portraying a particular characteristic deviates from 50%. The error about the estimate of the mean budget within the same parameters described in the text above was 1.7% (90% confidence) and 2.1% (95% confidence). See appendix I.

- 3.18 The quality of the data provided in response to any survey primarily reflects the questionnaire design, and the familiarity of institutional management information systems with the information sought. The work undertaken by the national coordinators to present the questionnaire properly has helped to mitigate many of the difficulties associated with what most agree was an extensive and in part complex form of enquiry. Nevertheless, considerable time has been devoted to cleaning the data and to the process of scrutinising the data.
- 3.19 Despite the internal guidance and checks embedded in the questionnaire, many respondents have provided textual answers to empirical questions. In some countries, the textual, rather than the spreadsheet form of the questionnaire was used. Further translation of the responses made in 14 different languages¹² has largely been undertaken mechanically: we generated tables so that the computer routines could 'look up' the appropriate translations for the standard response categories, and covert financial data to euros. While currency exchange rates have been particularly volatile in recent months, we have used rates at September 2008¹³, as these were prevailing during the main period of the survey canvass.
- 3.20 Some further adjustment has been necessary to the raw data supplied by respondents to take account of implausible information (e.g. more staff devoted to digitisation work than were employed in the institution as a whole). We regarded how institutions classified themselves, in particular. Several multi-purpose institutions failed to report themselves under the domain of their principal activity, and we adjusted the database for summary classification purposes to account for this. For instance, one particular national library was principally a library although its collections included materials otherwise held by archives and museums, and it reported itself as 'another type of institution'. All adjustments made by ourselves, or national coordinators scrutinising the data, have been noted on the database. Inevitably some further inconsistencies in the data will have gone undetected despite these rigorous checks, but in our opinion these represent an insignificant proportion of the database.

¹² Czech; Dutch; English; Estonian; French; German; Hungarian; Latvian; Lithuanian; Polish; Portuguese; Romanian; Slovenian; Spanish.

¹³ See appendix I.

3.21 The proportion of all responders providing valid data is shown in the blue cells, according to each question in the following three tables.

Table 7 Questionnaire completion rates, extract 1

4 Does your institution provide a national or regional service?	94%	6			
8 Total number of staff in your institution (full-time equivalents)	91%	10	90%	11	Currency:
9 Annual revenue budget for your organisation (for the whole institution)	76%	12	81%	13	100%
10 Full-time equivalent number of staff managing or engaged in digitisation	84%	15	82%	16	
11 i Does your institution have a written digitisation plan?	94%	17			
ii If 'yes', does this cover the long-term preservation of digital items?	22%	18	Year:		
12 What is your institution's annual budget for digitisation (in euros)?	66%	19	70%	20	
13 What percentage of your digitisation costs to date were funded by:					
i your own institution's resources?	72%	21			
ii government programmes?	72%	22			
iii private donations?	72%	23			
iv other 'in kind' support? (e.g. value of partnership help)	72%	24			
(Total)	72%	25			
14 What percentage of your analogue collection(s):					
i does NOT NEED to be digitised?	78%	26			
ii has ALREADY BEEN digitised?	78%	27			
iii STILL NEEDS TO BE digitised?	78%	28			
(Total)	78%	29			
15 i Do you possess an online catalogue of your collection(s)	85%	30			
ii If 'yes', does this distinguish digitised materials?	36%	31			
16 What percentage of your digitisation work was carried out:					
i in-house?	83%	32			
ii by an external contractor?	83%	33			
iii by a partner institution?	83%	34			
(Total)	83%	35			

^{3.22} The proportion of all responders providing valid data is shown in the blue cells and the number in the yellow highlights shows the cell to which the specific answer has been coded in the database.

3.23 As shown in Table 7, the response to the contextual questions was generally good, and with the exception of the digitisation budget, most institutions were able to provide resourcing information. More than two-thirds of responders were able to provide comprehensive information about their sources of funding for digitisation work, and more than four-fifths provided details about the organisation undertaking the work. The rate of progress made towards the digitisation of collections was reported by more than three-quarters of respondents.

Table 8 Questionnaire completion rates, extract 2

Digitisation of pictorial images (photographs, paintings, etc.)		pictorial images alone:			
17	Resolution in dots per inch (dpi)	54%	36		
18	File format of master files	57%	37		
	(Specified other)	5%	38		
19	What is the average cost per image (in euros) achieved?	25%	39		
Digitisation of text: (include manuscript, print and microfilm)		text with images:		text only:	
20	Resolution in dots per inch (dpi)	41%	40	24%	41
21	Most frequent file format of master files:	45%	42	28%	43
	(Specified other)	3%	44	4%	45
22	What percentage of digital output is OCRed?	37%	46	19%	47
23	What is the average cost per page (in euros) for OCR achieved? (i.e. inclusive of digitisation cost)	12%	48	7%	49
24	How many digitised 'image' pages remain to be processed by OCR, to provide searchable user access?	11%	50		
Digitisation of audio material:		audio materials:			
25	File format of master files	24%	51		
	(Specified other)	4%	52		
26	What is the average cost per hour (in euros) achieved?	8%	53		
Digitisation of video and film:		film:		video:	
27	File format of master files	19%	54	13%	55
	(Specified other)	9%	56	7%	57
28	What is the average cost per hour (in euros) achieved?	7%	58	4%	59
29	Do you use simultaneous high-speed A-V digitisations equipment?	23%	60		
Usage of and access to digitised materials:					
30	Is access mainly to full service free, paid or restricted?	82%	61		
	(Specified other)	14%	62		
Proportion of digitised materials available on the internet:					
31	What percentage of the material you have digitised is publicly available on the internet?	69%	63		
32	How many user requests to access DIGITISED materials were there last year:				
i	Online via the internet?	32%	64		
ii	By any other method? (e.g. CD-Rom within the institution)	34%	65		

The proportion of all responders providing valid data is shown in the blue cells and, the number in the yellow highlight shows the cell to which the specific answer has been coded in the database. The question number precedes the written question on the representation of the questionnaire

3.25 Table 8 sets out another 'high-level' part of the questionnaire which sought information about the unit costs of digitisation, and the formats adopted in order to establish some sort of qualitative measure for how these might differ. It should be born in mind that

the questionnaire was generic to all types of institution. Therefore, a low response to a particular question may simply reflect the fact that the question is not relevant to all types of institution. In this context, a substantial variation in response can be explained by the different types of material held by specific types of institution.

3.26 However, this part of the survey also explored the access policies adopted by each institution, to which a relatively high proportion of respondents were capable of providing answers. Yet these same responders evidently found it more difficult to provide information about the level of user access, and this was also reflected in the variation in values reported, as we describe later in this report.

Table 9 Questionnaire completion rates, extract 3

	Analogue collection		Digital output									
	Unit of measurement:	Total units:	Digitised units:	Planned digitisation		Pages:	Cost (euros):					
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]				
Objects/items/materials in collections:												
a Archival records	Please choose unit of measurement: 66	99%	33%	67	23%	68	15%	69	7%	70	8%	71
b Rare books	Volumes	34%	72	25%	73	14%	74	8%	75	8%	76	
c Other books	Volumes	49%	77	30%	78	19%	79	9%	80	8%	81	
d Newspapers	Issues	23%	82	18%	83	10%	84	6%	85	6%	86	
e Serials	Volumes	34%	87	22%	88	11%	89	7%	90	6%	91	
f Manuscripts	Number	31%	92	24%	93	12%	94	8%	95	7%	96	
g Sheet music	Number	18%	97	12%	98	6%	99	4%	100	4%	101	
h Microforms/microfilms (not included elsewhere)	Number	20%	102	9%	103	5%	104	3%	105	3%	106	
i Maps	Number	33%	107	23%	108	13%	109			5%	110	
j Photographs	Number	51%	111	43%	112	25%	113			11%	114	
k Engravings/prints	Number	32%	115	25%	116	12%	117			5%	118	
l Drawings	Number	31%	119	25%	120	13%	121			6%	122	
m Posters	Number	27%	123	20%	124	10%	125			4%	126	
n Postcards	Number	26%	127	19%	128	9%	129			3%	130	
o Paintings	Number	32%	131	25%	132	12%	133			5%	134	
p Any other two dimensional object not shown above	Number	18%	135	14%	136	8%	137			3%	138	
q Three dimensional works of art	Objects	25%	139	20%	140	10%	141			4%	142	
r Man-made artefacts	Objects	19%	143	17%	144	7%	145			4%	146	
s Natural world specimens	Objects	12%	147	10%	148	5%	149			2%	150	
t Other objects in collections	Objects	16%	151	13%	152	7%	153			2%	154	
u Monuments (Complete supplementary form if you use this line)	Number	8%	155	6%	156	3%	157			2%	158	
v Film	Hours	24%	159	17%	160	9%	161			4%	162	
w Video recordings	Hours	25%	163	16%	164	10%	165			4%	166	
x Audio (music and other recorded sound)	Hours	27%	167	18%	168	11%	169			5%	170	
y Other items not classified above	Number	12%	171	9%	172	5%	173			2%	174	
z Average pages per metre of archive shelving	Pages/Mtr:	9%	175									

The proportion of all responders providing valid data is shown in the blue cells and the number in the yellow highlight shows the cell to which the specific answer has been coded in the database.

3.27 Apart from the detailed information requested about individual projects (which we report on below), the most detailed section of the questionnaire traced the digitisation undertaken compared with the types of analogue materials held in the institutions' collections. The completion rate partly reflects the absence of particular types of material within the broad mix of institutional types responding.

Furthermore, the difference in completion rates between analogue and digitised collections is, in part, explained by the lack of digitisation activity in a number of institutions holding such materials.

- 3.28 However, it will also be observed that the rate of completion diminishes, from left to right, for the cells on all rows of this section of the questionnaire. Therefore, although some extremely valuable statistics have derived from this part of the survey, it is evident that the results are based upon the experiences of a lesser number of institutions capable of researching their activities to this level of detail.
- 3.29 Another consideration relating to the survey response relates to the 'soft' information requested about the number and types of digitisation collections / projects each institution had been undertaken. Non-completion of this part of the survey arises either because the institution had no significant progress to report, or the information requested was a 'task too onerous' for the respondent to record on the questionnaire. Therefore, although we have analysed the responses, the results do not throw any further light on the quality of the overall statistical part of the survey.

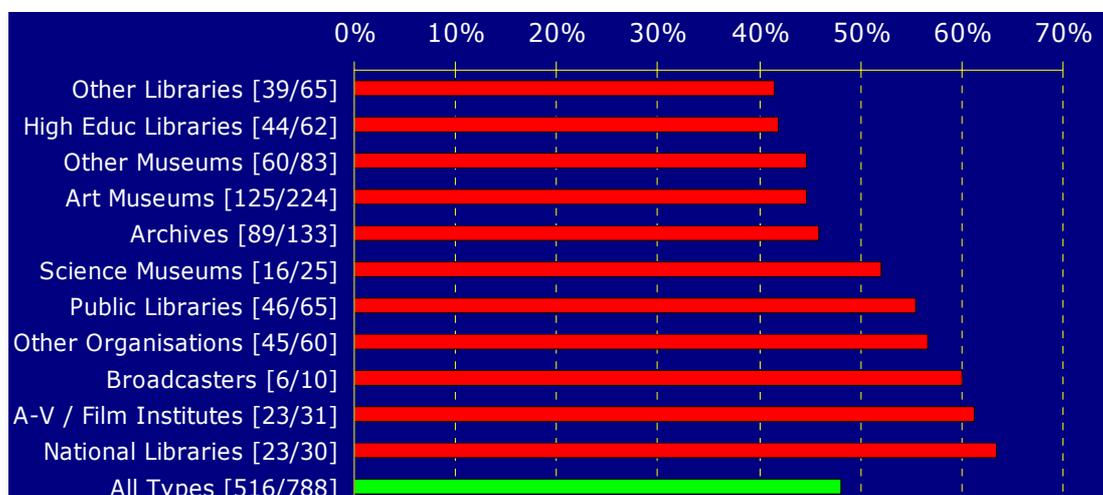
SURVEY RESULTS

- 3.30 A fuller analysis of the survey including summary tables for each survey question is set out in the appendices, where the analyses are classified according to the main type of cultural institution. Several conclusions arise from this analysis.

Budgets devoted to digitisation, and the proportion of the total budget of the institution this represents

- 3.31 Half the survey respondents (52%) either supplied no information about the amount set aside in institutional budgets for digitisation, or declared that no resources were earmarked for the purpose. The remainder of institutions surveyed (i.e. those possessing a digitisation budget with an earmarked amount to spend) was distributed as follows.

Figure 3 Institutions possessing a specific budget for digitisation activity



[The first figure in brackets indicates the number of institutions responding to the question. In this instance, the figure includes those who responded saying they had a budget heading, although it contained no resources. The bars in the diagram only represent the institutions with allocated resources.]

- 3.32 This sum was the amount identified in the institution’s own annual budget to provide for digitisation work. Respondents were asked to include the cost of the staff time devoted to digitisation projects in this estimate, but to exclude external funding, if this was not accounted for by the institution. However, it became evident that few institutions included the cost of staff time in this budget, and to achieve a comprehensive total we have adjusted the reported total, later in this report.
- 3.33 Based upon exchange rates in September 2008, the unadjusted total value of the digitisation budgets for responders was €80 million, representing 1.1% of the total reported institutional budgets (€7 billion). This latter budget was defined as the approved revenue planned spend, including the annual provision for staff and running costs of the whole institution, the cost of planned purchases to replace items in the collections, depreciation, maintenance of IT, and other such recurring institutional requirements.
- 3.34 However, not all respondents provided data for both budgets headings, and on a like-for-like basis the proportion was 0.88% at the mid-point of the distribution of 441 responding institutions. On the basis of projecting the survey results to all relevant institutions, we have estimated the proportion to be 0.71%. For most institutions, the formal budget identified for digitisation is a small proportion of the total day-to-day expenditure of the organisation.

To put this proportion in context, it should be noted that 9.5% of all responders had undertaken no digitisation activity to date¹⁴.

Table 10 Financial resources identified in budgets

Type:	Total for:	Institution	Digitisation	%
		€ millions [1]	€ millions [2]	
Archive/records office		223.0	10.0	4.5
Audio-visual or film institute		150.1	3.1	2.1
Broadcasting institute		2,578.9	6.6	0.3
Art museum (archæo/hist)		1,087.8	6.4	0.6
Science tech museum		171.3	2.1	1.2
Other type of museum		294.0	2.8	0.9
National library		676.0	25.0	3.7
Higher education library		517.5	2.0	0.4
Public library		250.1	2.0	0.8
Special or other library		530.9	3.9	0.7
Other organisations		543.1	16.2	3.0
	All types	7,022.5	80.0	1.1

3.35 Earmarked sums to undertake digitisation projects make up a larger proportion of institutional budgets in archives and national libraries, perhaps reflecting the concentration of materials in these institutions.

Full-time equivalent staff time devoted to digitisation work

3.36 The definition of ‘equivalent full-time staff’ specified those persons employed by the institution, assuming part-time staff were converted to the equivalent number for a full working week. This included both permanent and temporary staff on the establishment, but excluded contractors’ staff. Within this total, the time of the institution’s own staff engaged on digitisation projects – including planning and managing in-house and contracted projects, preparing and digitising materials and enhancing digitised output to widen accessibility – were to be accounted for as ‘digitisation staff’.

3.37 The majority of responders did not identify the cost of staff time in their budgets, and a fairer representation of the true investment in digitisation activity would need to include this element.

¹⁴ Based on those institutions marking their questionnaire ‘None’, and a scrutiny of the responses made to individual questions by others.

3.38 The relative distribution of staff time across the types of cultural institution appears more even than the budget allocations. This is shown in Table 11, although it should be noted that broadcasting organisations are much larger entities, dwarfing their technical digitisation units. If (as shown in Table 10) the identified digitisation budgets represented 1.1% of the total, it would be reasonable to assume that an additional 2.9% of the cost of the organisations’ staff time should be added to the budget allocation, i.e. a further €202 million within the survey institutions, providing a total estimated expenditure equivalent of €282 million.

3.39 Table 11 Full-time equivalent staffing resources

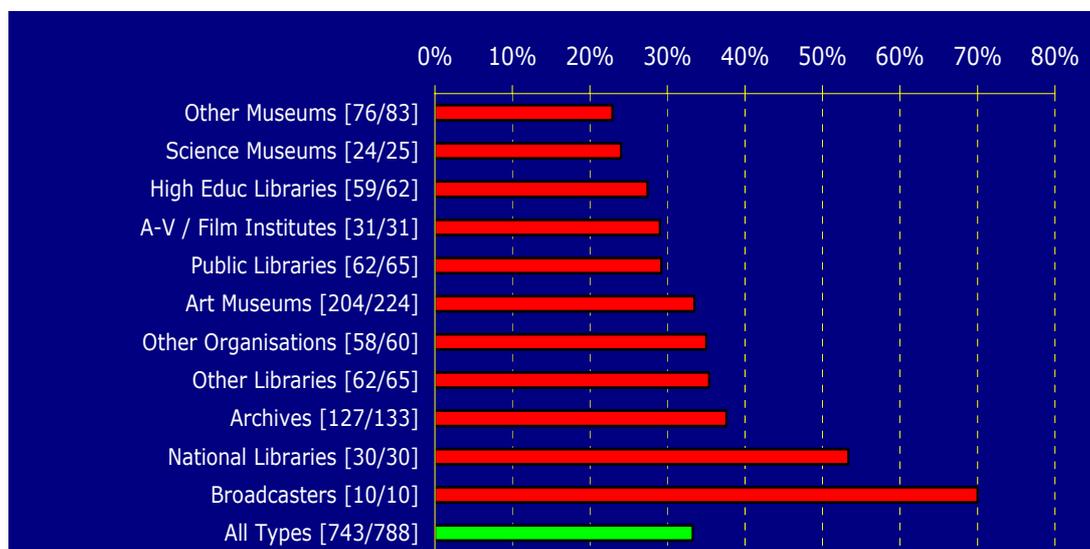
Type:	Total for:	Institution	Digitisation	%
		N ^o	N ^o	
		[1]	[2]	[2] / [3]
Archive/records office		6,303	434	6.9
Audio-visual or film institute		1,592	98	6.1
Broadcasting institute		16,378	127	0.8
Art museum (archæo/hist)		18,432	533	2.9
Science tech museum		2,860	58	2.0
Other type of museum		5,216	240	4.6
National library		11,152	400	3.6
Higher education library		7,843	184	2.3
Public library		7,191	106	1.5
Special or other library		2,730	82	3.0
Other organisations		2,506	104	4.2
	All types	82,203	2,366	2.9

Proportion of institutions that have adopted formal digitisation plans

3.40 About one-third of responding institutions (33%) had formulated a written digitisation strategy or plan. These were defined as any written strategy document to plan the digitisation of materials held by the institution for any period up to 2012. The main types of institution that have adopted such plans were broadcasters (70%) and national libraries (53%).

3.41 The overall pattern of digitisation planning is shown in Figure 4.

Figure 4 Formulation of a written digitisation plan/strategy



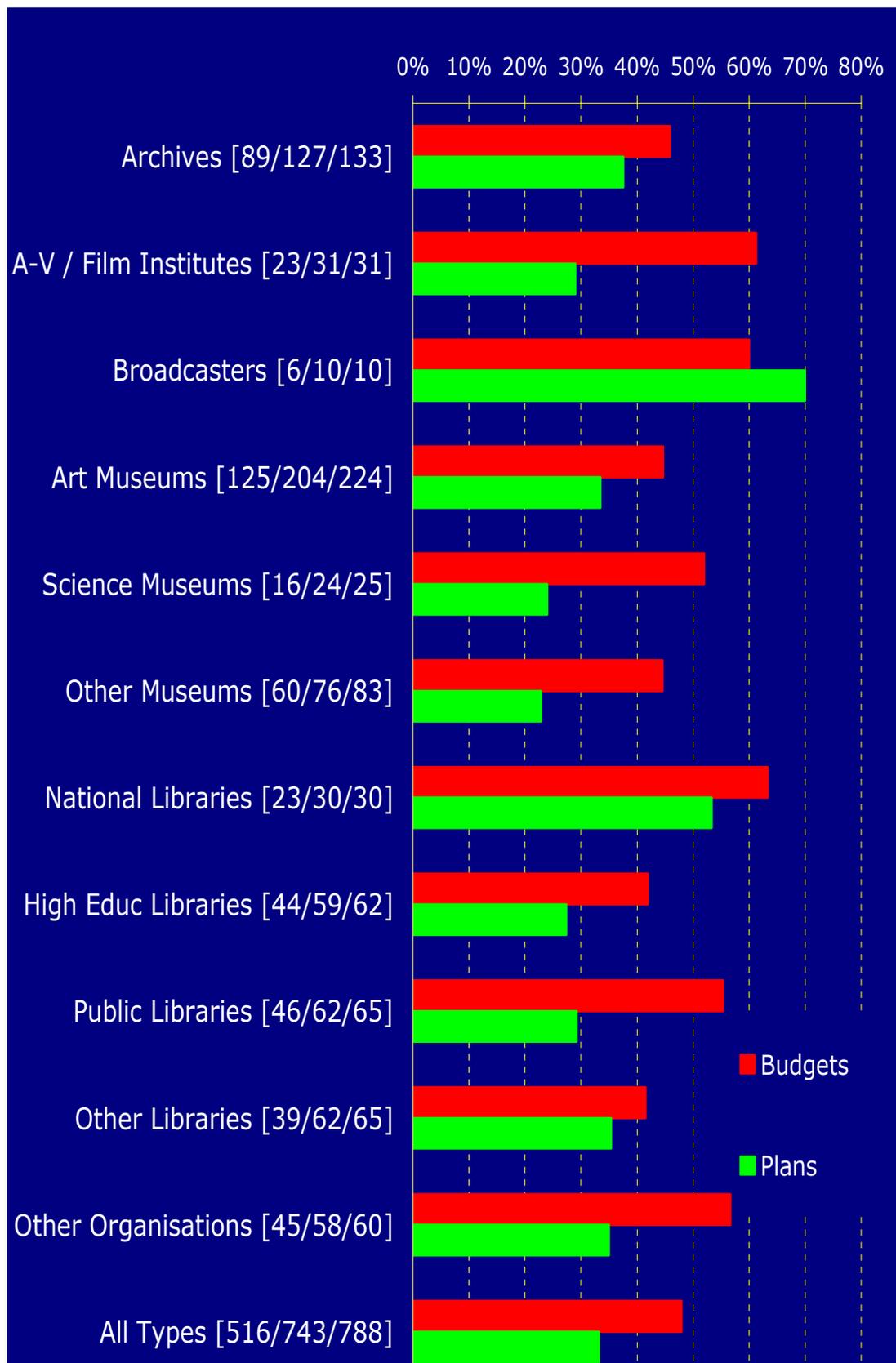
[The first figure in brackets indicates the number of institutions responding to the question; the second is the total number of survey respondents. Some will not have indicated that they possess a plan; the proportion that did is indicated by the bars in the chart.]

3.42 Of the 262 survey responders who had formulated digitisation plans, 150 (57%) confirmed that these included considerations for the long term preservation of the digitised assets. Discounting the small number of science and technology museums, the archives and national libraries appear to place more emphasis on plans for the long-term preservation of digital items.

Table 12 Digitisation plans

Type:	Total that:	Possess	% of all	Plan covers	% of those
		plans	institutions	long-term preservation	with plans
		N ^o	%	N ^o	%
		[1]	[2]	[3]	[3] / [1]
Archive/records office		50	38	37	74
Audio-visual or film institute		9	29	4	44
Broadcasting institute		7	70	4	57
Art museum (archæo/hist)		75	33	42	56
Science tech museum		6	24	4	67
Other type of museum		19	23	3	16
National library		16	53	10	63
Higher education library		17	27	10	59
Public library		19	29	11	58
Special or other library		23	35	12	52
Other organisations		21	35	13	62
All types		262	33	150	57

Figure 5 Resourcing of budgets and formulation of plans



[The first two figures in brackets indicate the number of institutions respectively responding to the question about their (1) BUDGET and their possession of a (2) PLAN; the third figure is the total number of (3) survey respondents. Some will not have indicated that they possess a budget or a plan; the proportion that did is indicated by the bars in the chart.]

Sources of funding for digitisation undertaken to date

3.43 About €30 in every €100 spent on digitising materials was funded under specific government programmes. €62 was found from the institutions’ own resources, and the remainder (€8) was met from other means (commercial activities, in-kind support and private donations). The distribution by type of institution is set out in Table 13.

Table 13 Sources of funding for digitisation activities

	Institute’s own resources	Government programmes ¹⁵	Private donations	Other support ¹⁶
	% [1]	% [2]	% [3]	% [4]
Archive/records office	43.1	33.0	17.3	6.6
Audio-visual or film institute	51.5	31.2	17.1	0.2
Broadcasting institute	94.4	4.4	0.0	1.1
Art museum (archæo/hist)	67.4	26.3	3.3	3.0
Science tech museum	73.7	15.8	10.1	0.4
Other type of museum	48.7	42.3	0.4	8.7
National library	44.6	36.0	16.6	2.8
Higher education library	39.6	30.7	6.1	23.5
Public library	67.8	21.5	7.4	3.3
Special or other library	53.3	41.1	1.0	4.6
Other organisations	63.5	26.0	2.5	8.0
All types	62.1	29.9	3.6	4.5

Who has undertaken the digitisation work, expressed as a proportion of all work completed to date

3.44 About one-third of all digitisation work appears to have been undertaken by external contractors. The majority of the work has been handled in-house (62%), although in a number of institutional types there appears to have been a significant amount undertaken by lead partner institutions; notably among archives, film institutes and national libraries. Further information about the use of contractors and partners is mapped in appendix IV (and described under *Digitisation partners* below.)

¹⁵ Government funding
Grants made by central, regional or municipal government for specific purposes, of which digitisation was an approved category.

¹⁶ Other support
This included other funding not categorised elsewhere, including ‘in kind’ help and revenues that derived from commercial arrangements.

Figure 6 Sources of digitisation funding

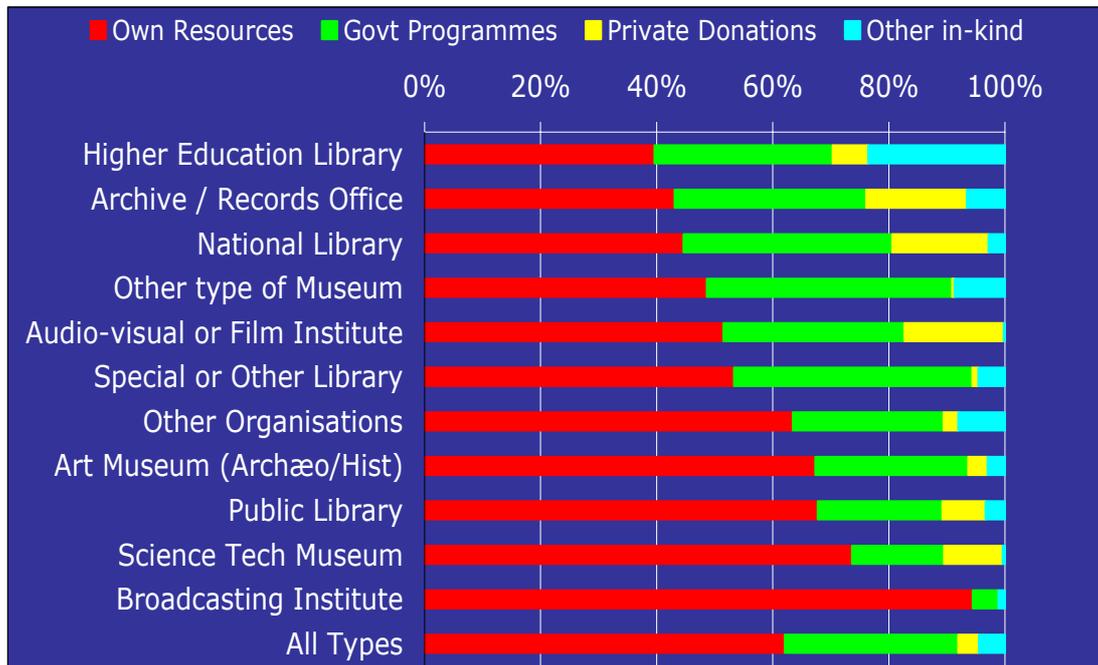


Figure 7 Undertakers of the digitisation completed to date

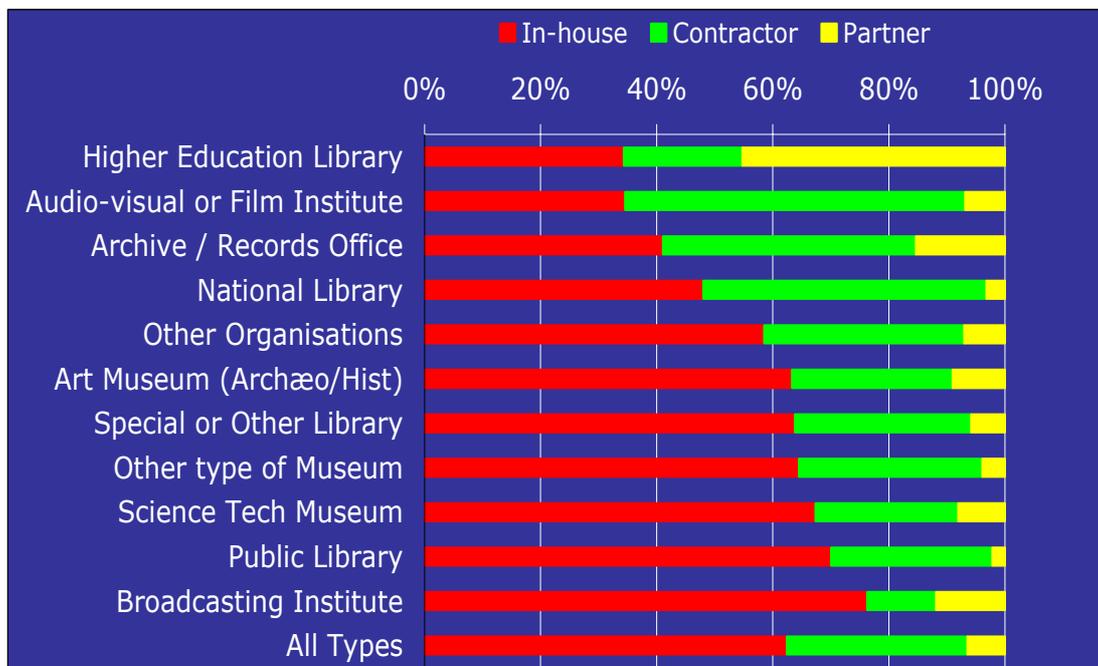


Table 14 Undertakers of digitisation projects

	In-house %	External contractors %	Partner institution %
	[1]	[2]	[3]
Archive/records office	41.1	43.6	15.3
Audio-visual or film institute	34.6	58.6	6.8
Broadcasting institute	76.3	11.9	11.9
Art museum (archæo/hist)	63.3	27.7	9.0
Science tech museum	67.4	24.6	8.0
Other type of museum	64.5	31.6	3.9
National library	48.1	48.8	3.2
Higher education library	34.3	20.5	45.2
Public library	70.1	27.8	2.2
Special or other library	63.8	30.4	5.8
Other organisations	58.5	34.5	7.0

Progress towards digitisation of institutional collections

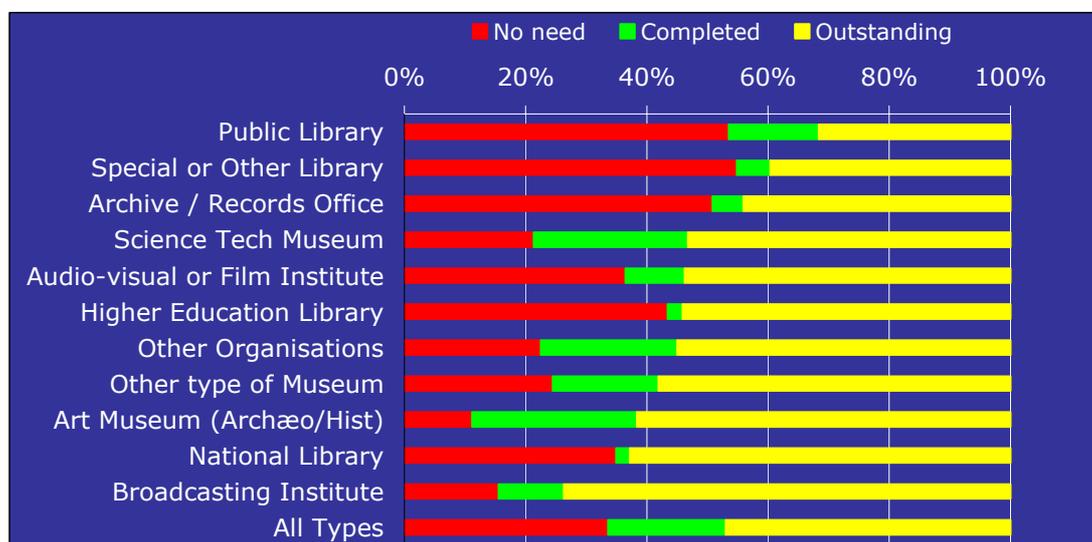
- 3.45 The simple proportion of the collection digitised can be a misleading statistic if regarded alone, since the collections held by different types of institution vary considerably in both the type of materials held and the size of their collections. For example, some institutions may contain a large number of items that are also held in other institutions, and therefore there is no need for all to duplicate the work. Instead, it is necessary to consider the progress in terms of the materials that institutions intend to digitise, and this objective varies considerably between different types of institution.
- 3.46 Therefore, it is necessary to exemplify the simple measure of progress for the whole of the cultural domain by regarding the progress made in individual types of institution, as set out in Table 15. This shows a more significant digitisation task ahead for national and relevant higher education libraries, in particular.
- 3.47 This measure of progress makes use of a notional ‘order book’ to represent only the proportion of collections subject to digitisation designs. Then the part of this notional order book that has been digitised can be compared with the remainder. Thus the proportion outstanding is equivalent to a ‘backlog’ of digitisation work still to be undertaken, which for national libraries was 28 times the volume so far completed.

Table 15 Progress made towards the digitisation of collections

Type of institution:	Part of collection digitised % [1]	'Order book' Completed % [2]	'Order book' Outstanding % [3]	Equivalent backlog [3]/[2] [4]
Archives	5.1	10.3	89.7	8.7
A-V or film institutes	9.8	15.4	84.6	5.5
Broadcasting institutes	10.8	12.8	87.2	6.8
Art/archæo museums	27.2	30.6	69.4	2.3
Science and tech museums	25.5	32.4	67.6	2.1
Other museums	17.5	23.1	76.9	3.3
National libraries	2.3	3.5	96.5	27.6
Higher education libraries	2.5	4.4	95.6	21.9
Public libraries	14.8	31.9	68.1	2.1
Special or other libraries	5.5	12.2	87.8	7.2
Other types of organisation	22.5	29.0	71.0	2.4

3.48 It has been suggested that the relatively high rate of progress made by museums may arise from the digitisation of the catalogue or documentation records of their collections, instead of the items or artefacts themselves.

Figure 8 Summarised progress by type of cultural institution



3.49 While the backlog (see Table 15) does not appear to be as daunting in the broadcasting institutes as it is in archives and the principal libraries, it is noticeable that the volume of material still remaining to digitise is significant. The following table compares the progress between the national and the local/regional institutions – there is no significant difference.

Table 16 Progress by the national prominence of certain types of institution

<i>The statistics in this table have not been weighted.</i>		Institutional prominence:			
Type:	Part of collection:	National	Regional or local	Unknown	Total for the type
Archive/ records office	No need to digitise	35.6%	44.5%	63.3%	41.6%
	Digitisation completed	10.2%	9.0%	6.3%	9.4%
	Digitisation outstanding	54.2%	46.5%	30.3%	49.0%
	<i>Valid responders</i>	38	57	3	98
Audio-visual or film institute	No need to digitise	32.8%	53.3%	...	36.9%
	Digitisation completed	10.1%	7.0%	...	9.5%
	Digitisation outstanding	57.1%	39.8%	...	53.6%
	<i>Valid responders</i>	16	4	...	20
Broadcasting institute	No need to digitise	15.6%	15.6%
	Digitisation completed	10.8%	10.8%
	Digitisation outstanding	73.7%	73.7%
	<i>Valid responders</i>	9	9
Museum of art, archæology or history	No need to digitise	13.3%	10.5%	14.3%	12.4%
	Digitisation completed	31.2%	29.0%	28.1%	30.3%
	Digitisation outstanding	55.5%	60.5%	57.6%	57.3%
	<i>Valid responders</i>	116	67	7	190
Other type of museum	No need to digitise	14.0%	19.2%	3.0%	15.0%
	Digitisation completed	23.7%	30.5%	49.6%	27.8%
	Digitisation outstanding	62.3%	50.3%	47.4%	57.2%
	<i>Valid responders</i>	42	24	5	71
National library	No need to digitise	45.6%	45.6%
	Digitisation completed	3.0%	3.0%
	Digitisation outstanding	51.4%	51.4%
	<i>Valid responders</i>	26	26
Higher education library	No need to digitise	56.8%	67.0%	84.2%	64.7%
	Digitisation completed	5.7%	7.4%	2.8%	6.1%
	Digitisation outstanding	37.5%	25.7%	13.0%	29.2%
	<i>Valid responders</i>	18	19	5	42
Public library	No need to digitise	51.7%	57.1%	64.5%	56.8%
	Digitisation completed	0.7%	12.6%	0.1%	10.8%
	Digitisation outstanding	47.7%	30.3%	35.4%	32.4%
	<i>Valid responders</i>	6	46	2	54
Special or other type of library	No need to digitise	40.3%	57.3%	32.7%	43.9%
	Digitisation completed	8.8%	4.9%	30.7%	9.2%
	Digitisation outstanding	50.9%	37.9%	36.7%	46.9%
	<i>Valid responders</i>	35	12	3	50
Other type of organisation	No need to digitise	19.0%	10.4%	66.7%	19.9%
	Digitisation completed	24.7%	28.9%	3.3%	24.4%
	Digitisation outstanding	56.3%	60.7%	30.0%	55.7%
	<i>Valid responders</i>	19	13	3	35

Please note that the more significant progress amongst museums may simply arise from their recording of objects on digitised catalogues rather than the full digitisation of objects held in their collections. The data may not be strictly comparable as a result.

3.50 A more detailed section was included on the survey questionnaire, which attempted to relate the digitised output to the analogue collections held by each institution. However, not all respondents were able to complete this, and not all those that did provided correspondingly consistent information relative to the analogue and digitised base. For this reason, the detailed analysis is not

consistent with the statistics relating to progress reported above, which are based on a much more comprehensive response. Regardless, providing the results are regarded with circumspection, a broad analysis provides a useful insight into the relative progress being made towards the digitisation of the main different types of material, even though there is no measure of the proportion of the analogue base that it is intended eventually to digitise.

Table 17 Progress made towards the digitisation of materials

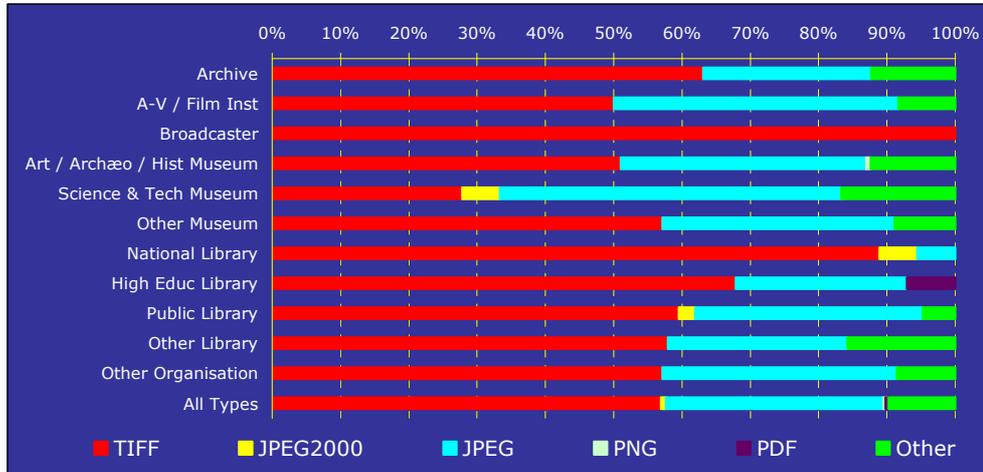
Type of material Held in collections	Measured in units: [1]	Analogue base Millions [2]	Insts. [3]	Digitised items Millions [4]	Insts. [5]	Progress % [6]	Adjusted % [7]
Archive records (1)	Metres	6.178	206	2.236	139	36.2	53.6
(2)	Volumes	2.538	6	0.010	5	0.4	0.5
(3)	Number	1,497.994	51	2.418	40	0.2	0.2
Books and serials	Volumes	223.523	383	4.061	240	1.8	2.9
Newspapers	Issues	248.238	269	16.734	171	6.7	10.6
Other printed matter	Number	146.326	244	5.213	192	3.6	4.5
Pictorial images	Number	177.909	403	10.170	342	5.7	6.7
Other physical objects	Number	207.342	196	9.175	154	4.4	5.6
Audio-visual materials	Hours	17.215	212	2.251	144	13.1	19.3
Unclassified other	Number	148.784	98	60.311	71	40.5	56.0
Mean progress (weighting all materials by the analogue collection) :						4.2	5.8

Note: "**Books and serials**" includes rare books; "**Other printed matter**" includes Manuscripts, Sheet music, and Microforms / Microfilms not counted under any other heading; "**Pictorial images**" includes Maps, Photographs, Engravings, Prints, Drawings, Posters, Postcards, Paintings and Other two-dimensional objects; "**Other physical objects**" includes Three-dimensional works of art, Man-made artefacts, Natural world specimens, Other items in collections and Monuments; "**Audio-visual materials**" includes Films, Video recordings and Audio materials.

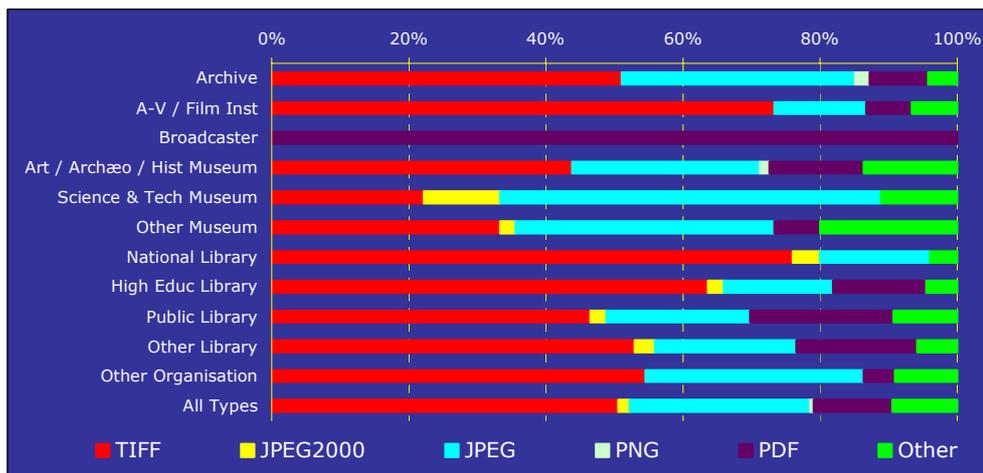
3.51 In Table 17, the relative progress measured under column [6] is based on the proportion that column [4] represents of column [2]. The rate of progress shown in column [7] is based upon an adjusted volume of digitised materials (column [4]) assuming an equivalent number of institutions reporting their analogue collections (column [3]). This may exaggerate the rate of progress because a proportion of the institutions that did not report the number of digitised units may have undertaken no digitisation of the kind. Therefore, for the institutions concerned, the actual rate of progress is most probably somewhere between the two figures in columns [6] and [7].

Figure 9 Comparative file formats employed for text and images

The most frequently adopted format for master files to digitise **PICTORIAL IMAGES**



The most frequently adopted format for master files to digitise **TEXT with IMAGES**



The most frequently adopted format for master files to digitise **TEXT ONLY**

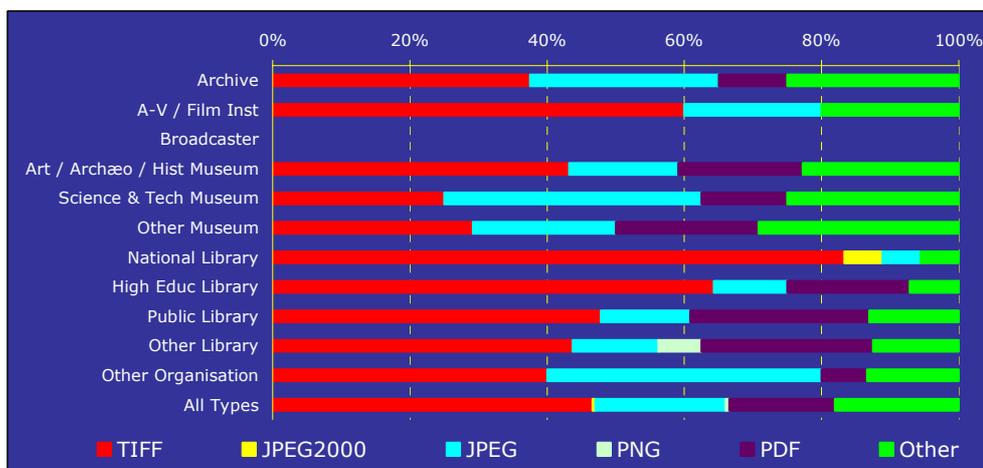
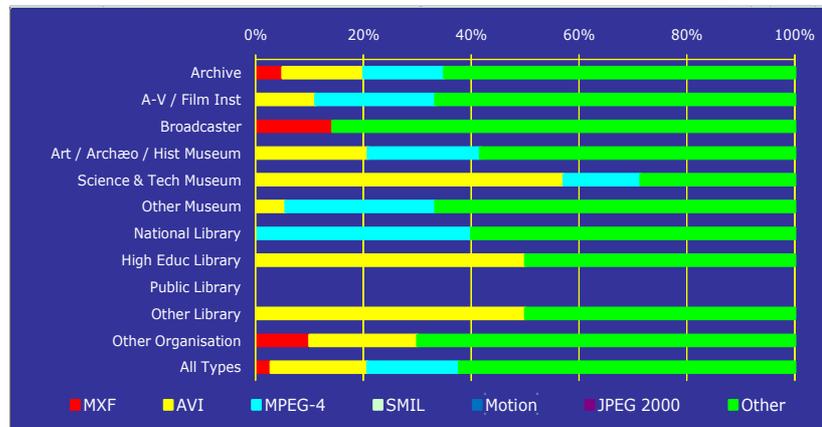
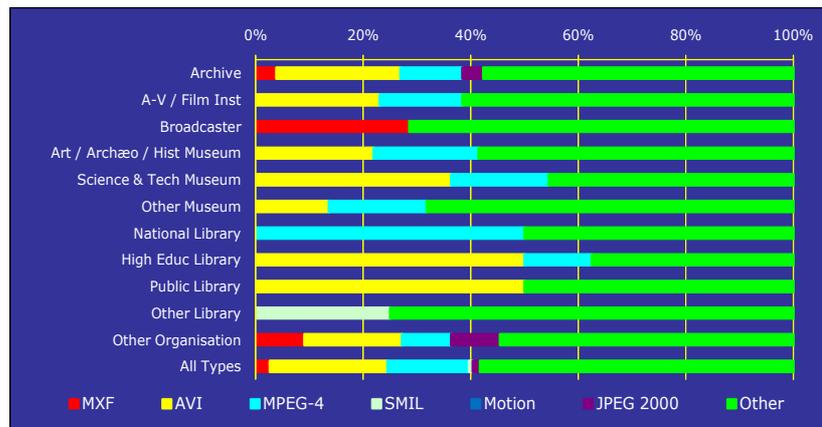
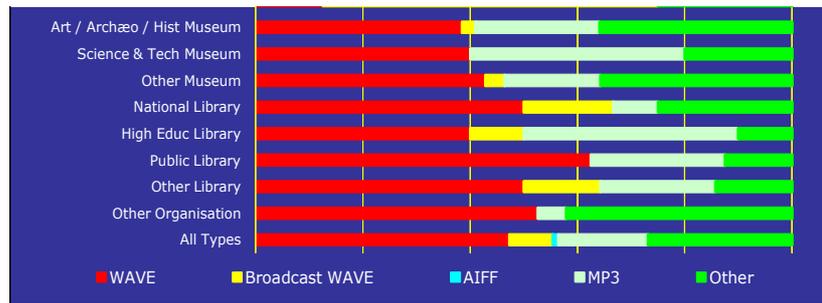


Figure 10 Comparative file formats employed for audio-visual materials



Most common formats for the master files relating to digitisation work undertaken

3.52 For digitised images and text, the most prevalent master format for files is based on the text interchange standard (*.TIFF), recorded at 300 dots per inch.

Table 18 File formats adopted for digitising images and text

	Images alone	Text and images	Text alone
Master format:	%	%	%
	[1]	[2]	[3]
TIFF	56.9	50.6	46.6
JPEG	31.9	26.3	19.0
PDF	0.4	11.5	15.4
Other	9.8	9.5	18.1
JPEG2000	0.7	1.7	0.5
PNG	0.2	0.6	0.5
All formats	100.0	100.0	100.0

3.53 In practice, the 'other' formats tended to be a combination of the main categories listed on the questionnaire. In this connection the combinations for images included mainly TIFF and JPEG, with references to PNG, DjVU and NEF, CR2 (photographs). For materials containing text and images, the 'other' combinations also included HTML, SGML, BMP, Excel, Word and basic TXT formats. For materials containing only text, the prevalent combination was TXT, HML, JPEG, and PDF.

3.54 A further consideration in the rate of progress towards digitisation relates to the volume of text materials scanned, but where the images are awaiting character recognition. Among survey responders, there were 41 million pages¹⁷ of scanned material still to be processed by optical character recognition equipment (OCR). We estimate that if these results are applied to all relevant institutions, then this current OCR task is equivalent to 494 million pages. While the ratio is higher among libraries, the mean proportion of materials subjected to OCR was 19% for text and images and 26% for text alone.

3.55 The digitisation of the materials described in Table 19 and Figure 10 evidently employs rather more diverse formats. Frequent references were made to DVD/CD for sound, but the formats for these were not stated. MXF and variants of MP/Mpeg and FLAC were also reported. For film and video, the most often quoted format related to BETA/VHS variants, DV and earlier versions of the MPEG standard.

¹⁷ Equivalent to 49% of the material converted into image format for those institutions that reported a proportion rather than the volume of material awaiting conversion.

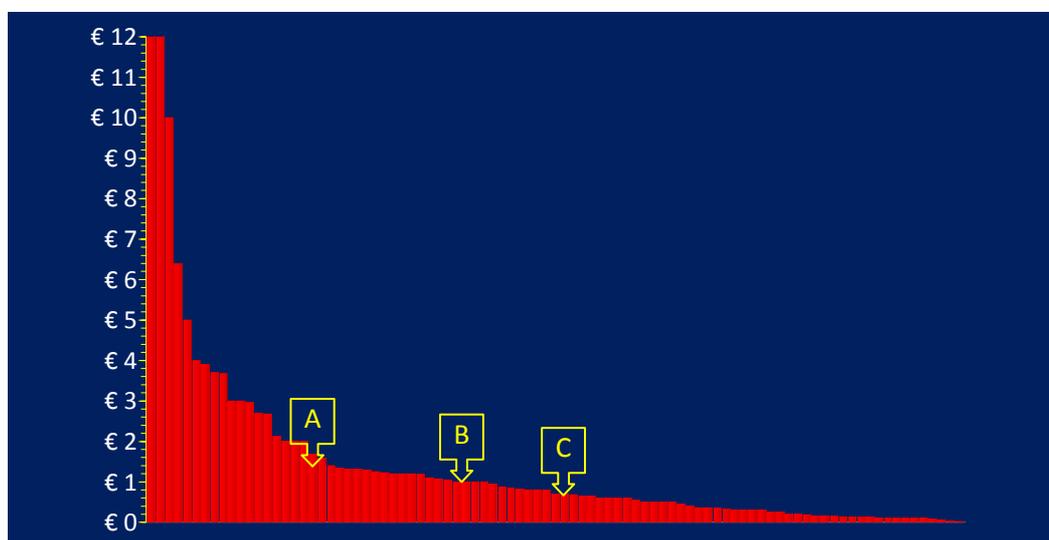
Table 19 File formats adopted for sound and motion images

Master format:	Audio	Master format:	Film	Video
	% [1]		% [2]	% [3]
WAVE	47.3	Other	58.3	62.3
Other	26.9	AVI	21.9	17.9
MP3	16.7	MPEG-4	15.2	17.0
Broadcast WAVE	8.1	MXF	2.6	2.8
AIFF	1.1	JPEG 2000	1.3	...
...	...	SMIL	0.7	...
...	...	Motion
All formats	100.0		100.0	100.0

Unit costs (per page/hour) of digitised materials

3.56 Two approaches were adopted in the survey to the measurement of digitisation costs: the cost per page/hour of materials already processed, and rather more detailed estimates of the cost applying to the planned processing of specific types of material (see *Cost of planned digitisation work according to type of material*, at 3.69).

Figure 11 Median cost of digitising text combining images on the same page



- € 1.41 A: Mean Arithmetic mean (sum of values divided by number of values)
- € 1.00 B: Mode Most frequently occurring or repetitive value in the distribution
- € 0.69 C: Median The value in the middle of the ranked distribution

3.57 Considering different circumstances and form of the analogue materials being processed, the cost will vary considerably between projects even within the same institution. Therefore, it should be

understood that the average cost represents a wide distribution of values. Furthermore, the reported costs of digitisation varied considerably between institutions, in some part reflecting different processes in the digitisation tasks and technology. This is demonstrated in the distribution of average costs for processing materials combining text and images on the page, shown in figure 11.

3.58 Consequently, the values for the main types of institution, shown in Table 20, are only indicative of actual benchmarks that may be drawn widely about the average. Keeping this in mind, the median cost of digitising pictures in art museums (€5) was rather more than the cost of digitising images in national libraries (€1).

Table 20 Median reported digitisation costs

Type of cultural institution:	Images alone €/page [1]	Text & images €/page [2]	Text alone €/page [3]	Audio €/hour [4]	Film ¹⁸ €/hour [5]	Video €/hour [6]
Archive/records office	3.00	0.58	0.77	29.11	100.00	100.00
A-V or film institute	5.36	0.72	0.32	17.92	128.89	7.00
Broadcasting institute	78.84	678.10	100.00
Art museum (archæo/hist)	5.00	1.00	0.80	24.23	54.00	35.02
Science tech museum	1.50	0.25	0.45	16.71	16.37	13.02
Other type of museum	6.00	1.98	0.32	38.33	26.67	26.67
National library	1.00	0.50	0.48	58.42	16.76	55.00
Higher education library	0.72	0.81	0.10	6.42	8.00	5.46
Public library	0.96	0.65	0.70	9.70
Special or other library	1.03	0.59	0.19	35.00	1,040.00	120.00
Other organisation	2.00	0.87	0.55	39.00	22.22	35.13

3.59 Less than 0.9% of survey responders and only 20% of broadcasters reported the use of simultaneous high-speed audio-visual digitisation equipment. This was defined as conversion equipment that can digitise several originals simultaneously and at faster speeds than the original play-back time. Some museums and university libraries reported the use of this equipment, but no film institutes.

¹⁸ Six institutions reported digitisation costs for films in excess of €1,000 per hour. The €1,040 cost under column [5] is based on only one a library attached to a film institute.

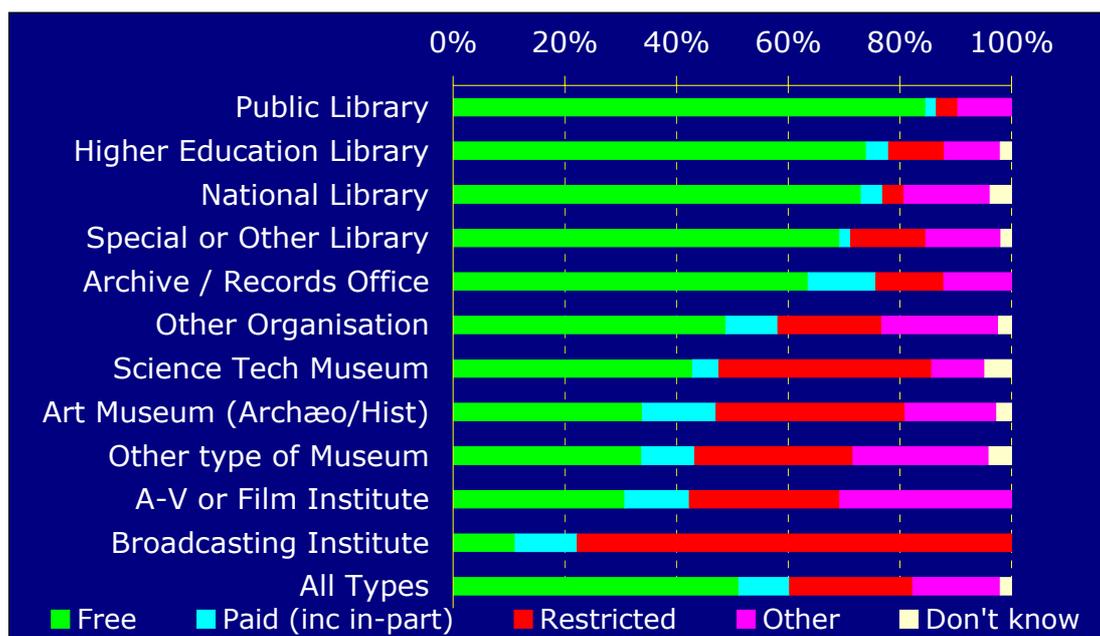
Access policies and the percentage of digitised materials publicly available on the internet

3.60 About one-half of all institutions and three-quarters of all libraries operate a free access policy to the materials they have digitised. In Table 21, the apparently large proportion of institutions adopting 'something else' arises because they were applying a combination of the distinct policy options contained on the questionnaire. Figure 12 provides a clearer impression of the access policies operating, and reflects the primary aims of the respective types of institution.

Table 21 Institutional access policies to their digitised materials

Type of institution:	Free to full service [1] %	Paid to full or part [2] %	Restricted (e.g. in-house) [3] %	Something else [4] %	Don't know [5] %	Valid insts [6] N°
Archive/records office	63	12	12	12	0.0	107
A-V or film institute	30	11	26	30	0.0	26
Broadcasting institute	11	11	77	0	0.0	9
Art museum (archæo/hist)	33	13	33	16	2.6	189
Science tech museum	42	4	38	9	4.8	21
Other type of museum	33	9	28	24	4.1	74
National library	73	3	3	15	3.8	26
Higher education library	74	4	10	10	2.0	50
Public library	84	1	3	9	0.0	52
Special or other library	69	1	13	13	1.9	52
Other organisation	48	9	18	20	2.3	43

Figure 12 Free access policies compared by type of institution



3.61 Assessing the number of beneficiaries gaining access to the materials has been more problematic, as evidenced by the lower completion rates to this part of the questionnaire (see Table 8).

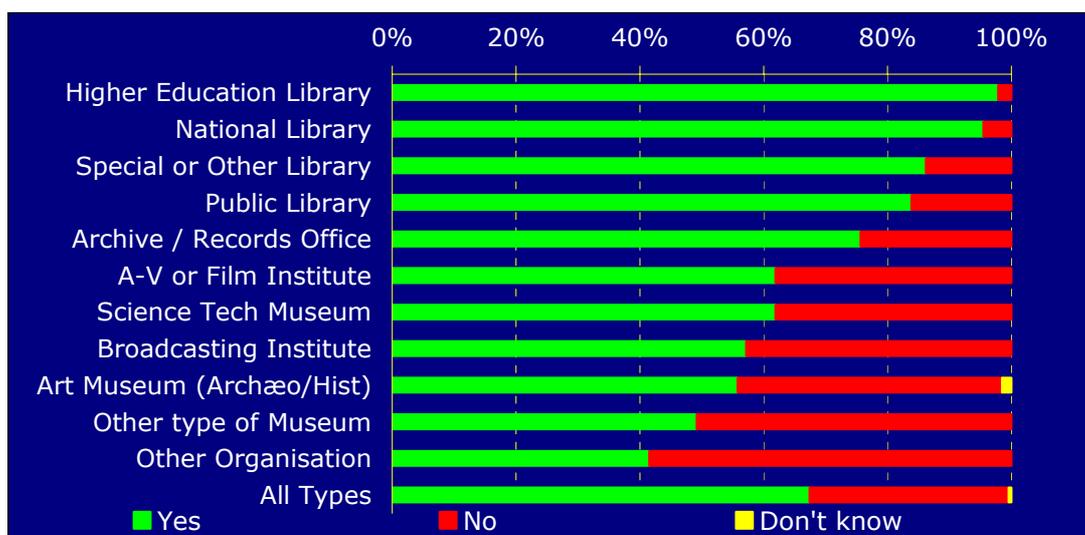
3.62 The first point to note is that not all institutions have an online catalogue of their analogue collections, and a considerably smaller proportion can distinguish the digitised content within this catalogue. Table 22 below is expressed as a proportion of only those responding to the question.

Table 22 Online catalogue reference to digitised materials

	Does your institution possess an online catalogue?				If 'yes', does it distinguish digitised materials?			
	Yes %	No %	? %	Resp N ^o	Yes %	No %	? %	Resp N ^o
Types of institution:	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Archive/records office	75.7	24.3	-	115	60.4	39.6	-	48
A-V or film institute	61.9	38.1	-	21	100.0	-	-	6
Broadcasting institute	57.1	42.9	-	7	75.0	25.0	-	4
Art museum (archæo/hist)	55.8	42.7	1.5	199	58.1	40.3	1.6	62
Science tech museum	61.9	38.1	-	21	62.5	37.5	-	8
Other type of museum	49.2	50.8	-	61	77.3	22.7	-	22
National library	95.5	4.5	-	22	88.9	11.1	-	18
Higher education library	97.9	2.1	-	47	75.9	24.1	-	29
Public library	83.9	16.1	-	62	41.0	59.0	-	39
Special or other library	86.2	13.8	-	58	53.6	46.4	-	28
Other organisation	41.5	58.5	-	53	57.1	42.9	-	14

3.63 Figure 13 clarifies the relative distribution and, as one would expect, shows that libraries have the most developed online cataloguing systems. Of those institutions that had online catalogues, three-quarters of national libraries (76%) and broadcasters (75%) distinguished their digitised materials.

Figure 13 Possession of online catalogues



3.64 The proportion of digitised materials that is publicly available on the internet varies considerably between institutions of all types. The median proportion for all institutions was 20%. The relative proportions reported by institutions belonging to each domain are set out in Table 23, which shows that libraries and archives have been most successful in widening access to their collections.

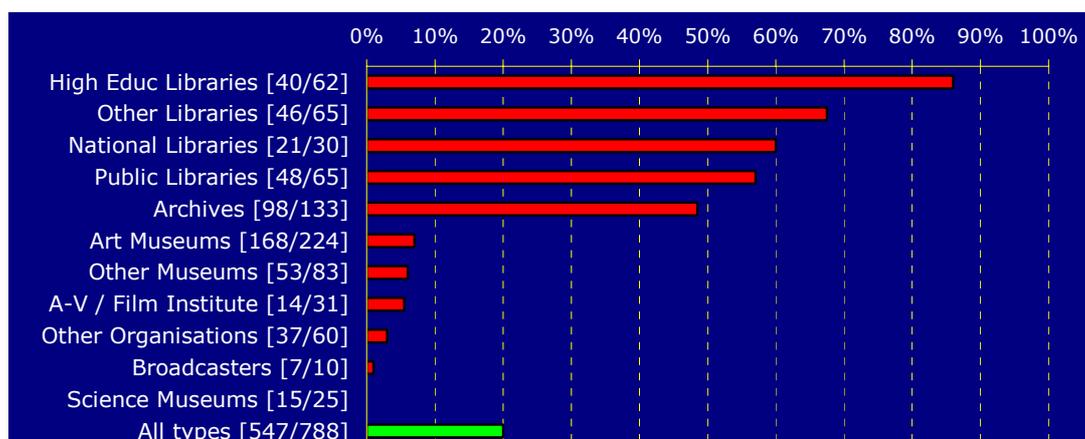
Table 23 Digitised materials publicly available on the internet

Domains:	Median, %
Archives	48.5
Audio-visual/broadcasters	4.0
Museums	5.5
Libraries	70.0
Others †	3.0
All domains	20.0

Note † Examples of other types of institution include technical centres for digitisation / ICT, Government Departments, agencies responsible for monuments, research and other heritage organisations.

3.65 A major factor relating to the availability of materials on the internet relates to rights. Whilst the function of most institutions is to impart information or provide reference material to all, the digitisation of materials in the audio-visual institutions can be more complicated. Whilst there is a focus on transferring deteriorating materials from obsolete to new carriers, there are significant restrictions on what can be made publicly available on the internet. The constraints arising from performance rights, and the existence of multiple holders of materials are considerable. Similarly in museums much activity may be devoted to the cataloguing or referencing of materials for internal purposes. Furthermore, many digitised materials may be restricted for viewing on the premises, as part of exhibitions.

Figure 14 Proportion of digitised materials available on the internet



[The first figure in brackets indicates the number of institutions responding to the question, and the second shows the total number of questionnaire responders. The bars in the diagram show the average (median) proportion of digitised materials that are available on the internet.]

3.66 The data provided on the number of online (internet) user requests to access digitised materials were particularly inconsistent, despite considerable time spent on validating the responses. Some data undoubtedly refer to site hits, and others to visitor episodes. An example of high volume of user access was reported in one country, where 90 million internet user requests related to five archives. However, upon investigation these figures appeared to be a count of the number of internet accesses. In another instance, an institutional response of 18 million user requests was reported as false, and that it was not possible to count user requests. The survey established that about three-quarters of all libraries operated a free access policy to digitised materials (see 3.60). The operation of restricted or paid access policies tended to be more significant amongst other types of institution, and the overall volume of internet requests for digitised materials needs to be regarded in this light.

3.67 Given the inconsistency in the data, the figures provided in the report of survey for the level of internet user requests to digitised materials made over the internet need to be treated with extreme scepticism.

3.68 There was rather less variation in the survey figures about the number of other forms of access to digitised materials. These refer to all other non-internet accesses made, for example, by CD-Rom within the institution. For this type of enquiry, it is evident that the national libraries provide the main point of on-site access to digitised materials, although archives also satisfied a significant demand among the survey respondents last year. (The response rate to the question is distributed similarly to the overall survey

response, providing some confidence in the pattern of off-line requests reported in the survey.)

Table 24 Off-line user requests to access digitised materials

Type of institution:	Survey sum of user requests <i>Distribution</i>		Responders to the question <i>Distribution</i>		Responders to the survey <i>Distribution</i>	
	N° [1]	% [2]	N° [3]	% [4]	N° [5]	% [6]
Archives	1,583,971	5.82	34	13	133	17
A-V/film institute	11,896	0.04	11	4	31	4
Broadcasters	3,000	0.01	2	1	10	1
Art museums	378,550	1.39	88	33	224	28
Science museums	31,330	0.12	12	4	25	3
Other museums	8,350	0.03	40	15	83	11
National libraries	24,750,384	90.92	10	4	30	4
High educ. libraries	338,481	1.24	15	6	62	8
Public libraries	3,907	0.01	23	9	65	8
Other libraries	81,954	0.30	14	5	65	8
Other organisations	30,909	0.11	21	8	60	8
All types	27,222,732	100.00	270	100	788	100

Cost of planned digitisation work according to type of material

3.69 As described above, the estimated cost of plans for digitising specific materials was explored in the survey (as distinct from the cost of work already undertaken). The digitisation plan was defined as a written strategy document to cover the digitisation of materials held by the institution for any period up to 2012. Within this definition, the cost of planned digitisation was the equivalent of either the institution’s estimate of what a contractor would charge to digitise the materials, or the financial resources for digitisation identified in the plans. Whereas we have seen that the majority of institutions have not set formal digitisation plans, some institutions have reported their intentions, and shown the cost accordingly. The total cost of plans reported by respondents (for projects to 2012) was €408 million – more than five times the reported current annual budgets.

Table 25 Reported expenditure planned on digitising materials

Type of material:	Cost of plans			Measured in units of:
	€ millions	No. Inst	% of [1]	
	[1]	[2]	[3]	[4]
Archive records by: Metres	18.842	41	4.6	Metres
Number	0.471	7	0.1	Number
Rare books	65.638	56	16.1	Volumes
Other books	150.983	53	37.0	Volumes
Newspapers	21.679	34	5.3	Issues
Serials	41.289	26	10.1	Volumes
Manuscripts	8.325	46	2.0	Number
Sheet music	3.556	18	0.9	Number
Micro-forms/micro-films	0.658	10	0.2	Number
Maps	0.421	33	0.1	Number
Photographs	7.645	73	1.9	Number
Engravings/prints	0.711	30	0.2	Number
Drawings	0.781	36	0.2	Number
Posters	0.148	21	0.0	Number
Postcards	0.184	17	0.0	Number
Paintings	0.394	30	0.1	Number
Other 2D objects	0.367	17	0.1	Number
3D works of art	0.499	21	0.1	Objects
Man-made artefacts	8.209	19	2.0	Objects
Natural world specimens	28.28	8	6.9	Objects
Other items in collections	1.98	6	0.5	Objects
Monuments	2.513	4	0.6	Number
Films	22.382	17	5.5	Hours
Video recordings	9.908	22	2.4	Hours
Audio materials	4.882	32	1.2	Hours
Any other items	7.398	8	1.8	Number
Total cost of plans	408.144	184	100	

3.70 The emphasis in reported plans was placed upon the digitisation of printed materials; 70% of the investment was devoted to books, newspapers, serials and manuscripts.

3.71 The other significant planned investments include the digitisation of films and videos (8% of the total – column [3]); man-made artefacts and natural world specimens (9%), and archive records (nearly 5%). The median cost of digitising the different types of material actually represents a considerable variation reported by individual institutions. (The number of institutions providing the data summarised in the table is shown in each of the columns headed 'inst'.) Further information concerning the average number of pages contained in the units that institutions plan to digitise is set out in the more detailed tables in the appendices.

UNIVERSAL ESTIMATES

Projection method

- 3.72 The survey results only relate to the institutions that participated in the canvass. However, the sample was designed so that, as far as possible, the results could be reasonably attributed to all the other relevant cultural institutions in each country, providing, of course, there were sufficient responders upon which to base reliable estimates.
- 3.73 We have previously described how the summary statistics arising from the survey have been weighted to better reflect the actual distribution of institutions in each country and for the EU27 area as a whole.¹⁹ The weighting of the survey data was conducted at the same time as projecting the estimates for the complete universe of institutions in each summary area. Therefore, the statistical tabulations, reproduced at Appendix I, show the simple (un-weighted) analysis of the survey results, side by side with the projected totals (where relevant) and the weighted summaries.
- 3.74 We tried several different approaches to projecting the survey results to estimate the digitisation activity for the complete population of relevant institutions. In each approach, we undertook the modelling for individual types of institution. For example, national libraries were treated as a discrete group, and their survey results were not applied to represent other types of library. The types coincide with those listed in previous tables, although we also treated each type of agency responsible for monuments separately. These were:
- * archive/records office;
 - * audio-visual or film institute;
 - * broadcasting institute;
 - * museum of art, archæology or history;
 - * museum of science and technology (or ethnology);
 - * other type of museum;
 - * national library;
 - * higher education library;
 - * public library;
 - * special or other type of library;
 - * cultural ministry – responsible for monuments;
 - * national agency – responsible for monuments;
 - * regional agency – responsible for monuments;

¹⁹ The method of weighting is described at 3.7 and in the following paragraphs.

- * municipality - responsible for monuments;
- * other type of organisation.

3.75 In our first approach, we attributed the mean value for responders of the particular type of institution to other institutions of the same type in the same country. However, this tended to exaggerate the estimates for two reasons: responders were probably more actively engaged in digitisation activity than other institutions, and the responders' survey values were not normally distributed, so the response from one very active institution would inflate the arithmetic mean for the remainder. Therefore, we opted to base the projections on median values for the type of institution (i.e. the point in the middle of the ranked distribution of values).

3.76 We call the method of projecting estimates to represent the whole population of relevant institutions *grossing*. There were two stages:

1st order: The observed survey measure (median) for the responders of the specific type of institution in the specific area was attributed to the remaining relevant institutions of that type in the same area. The total for the type in that area, therefore, equalled the sum of the survey responders' values, plus the grossed results for the remainder.

2nd order: For those institutions of a particular type, where there were no representative survey responders of that type in that area, then the EU27 survey median value for that type of institution was employed.

3.77 In the instances where the grossing at the second order was undertaken, we consider the potential for distortion to have been minimised by adopting the survey median as the proxy for estimation, for the reasons explained at 3.75. Given the wide measure of dispersion in all the survey statistics, the approach adopted remains the safest in the circumstances.

3.78 A further refinement was incorporated within the second order of grossing to reflect the fact that not all institutions of the particular type were engaged in all forms of digitisation activity. The grossing formula included a weighting for the proportion of survey responding institutions providing a response to the particular activity being estimated. In summary, the grossing steps were undertaken as follows:

1st order: For each spatial area **with** valid survey data for the particular type of institution:

	(All relevant institutions – survey responders) x area survey value
+	Total survey value for responders
=	Grossed sum

2nd order: For each area **without** valid survey data for the particular type of institution:

	Providing relevant institutions of the type exist in the area concerned: (EU27 question responders / EU27 survey responders)
x	(EU27 survey value x relevant institutions of the type in the area)
=	Grossed sum

3.79 It is accepted that the second order of grossing produces stereotype national results, but these estimates are primarily required to provide aggregated EU27 figures (i.e. those areas [countries] with a similar number of relevant institutions of a particular type will show similar second order grossed results; these estimates are simply necessary to scale realistic international statistics). We undertook a final test of the overall method by examining the plausibility of the derived ratios calculated from the grossed statistics.

3.80 The grossing model that was used also undertook the weighting of the mean survey results (as described under *Adjusting for the pattern of response*, above), so that both the summed survey values and estimates based on the observed average reflect the distribution of all relevant institutions in each spatial area (country).

Projected estimates

3.81 The projected estimates arising from the survey results provide a base upon which the future progress in the cultural digitisation work can be reviewed. Therefore, in the following section we concentrate on estimating the size of the total ‘digitisation economy’ within cultural institutions (current and planned), and the access gained to the digitised content being generated. The financial resources assigned and the proportion of institutional staff time devoted to digitisation work provide measures of input into this economy, and the absolute volume of requests to access the digitised information provides an indicator of output. These are static measures, and do not take account of the potential cost savings or benefits to users gained in completing the digitisation agenda at a faster rate. Even so, at this stage, we recognise that these are only estimates and, for this reason, we wish to distinguish these from the analysis of the survey results in the previous part of this report.

Inputs

3.82 It is estimated that the equivalent of about 15,750 full-time staff²⁰ are devoted to either undertaking or supervising digitisation activities in the relevant cultural heritage institutions in Europe. The relative proportion of the whole staffing establishment assigned to digitisation work varies according to the scale of enterprise of the different types of institution. In the larger broadcasting and higher education institutes, the estimated proportion is under 1%, while in many of the smaller organisations, such as archives, the estimated average proportion is nearly 9%.

Table 26 Estimated staff resources of relevant institutions

Type of institution	Institutional No FTE [1]	Digitisation No FTE [2]	Proportion % [3]
Archive/records office	26,342	2,259	8.6
Audio-visual or film institute	8,310	341	4.1
Broadcasting institute	16,378	127	0.8
Art museum (archæo/hist)	94,502	5,835	6.2
Science tech museum	149,439	1,611	1.1
Other museum	31,391	2,161	6.9
National library	21,817	667	3.1
Higher education library	186,772	1,002	0.5
Public library	70,796	1,295	1.8
Special or other library	11,945	300	2.5
Other organisation	2,963	150	5.1
All types	620,655	15,748	2.5

3.83 As we have observed previously, a lesser proportion of all institutions²¹ (35%) have formulated a documented digitisation strategy or plan. Consequently, there is a lack of clarity about the intended future investment. Therefore, confining ourselves to the current situation, the total estimated annual value of budgets specifically identified by relevant EU cultural institutions for digitisation work was €261 million (at September 2008 exchange rates). Many institutions will have omitted the cost of staff time devoted to digitisation projects from their reported budgets, since such establishment expenses were usually held under other institutional budget headings. In this connection, the estimated proportion of the total institutional staff time devoted to digitisation work was 2.5%.

²⁰ Persons directly employed by the cultural institution on a whole-time basis, plus part-time staff converted to the equivalent number for a full working week. This included both permanent and temporary staff on the establishment, but excluded contractors' staff.

²¹ 35.3% of those answering the question, and 33.2% of all survey respondents.

3.84 Bearing in mind the considerable cross-over in purposes and collections between different types of institution, more than one-half of the estimated cultural digitisation budget (57%) would appear to arise from museums. However, this needs to be recognised in the context that museums represented 49% of the relevant institutions identified by National Coordinators. A bit more than one-fifth of the estimated digitisation budget (21%) would appear to be located in libraries, and a tenth (10%) in archives. The audio-visual institutions, including film archives and broadcasting institutions, also represent a significant sector of this digitisation economy, where their assigned resources equate to about 7% of the whole.

Figure 15 Estimated annual digitisation spend per head of population



Outcomes

3.85 The tangible results of the current investment made by institutions are equally hard to estimate with a great deal of certainty. The survey results alone summarise a complex background of institutional types within which the relative progress of each differs considerably. Table 19 (in the earlier part of this report) provides relative measures of the analogue base translated to digitised output (between 4.2% and 5.7% of collections to-date). Given the representation of the institutions making up this measure, this range of progress remains a reasonable assessment of current outcome.

3.86 We have previously described the difficulty of measuring the volume of internet requests to access the digitised output. Therefore, we

have concentrated solely on projecting the measures of 'off-line'²² requests which are more robust. These estimates reflect a reasonably uniform scale across those types of institution allowing public access to their materials, and are set out in the following table.

Table 27 Annual requests to access digitised materials²³

Type of institution (0.0 = too small to show to one decimal place)	On-site requests N° millions	Distribution %
	[1]	[2]
Archive/records office	2.432	5.7
Audio-visual or film institute	0.014	0.0
Broadcasting institute	0.003	0.0
Art museum (archæology/history)	3.564	8.3
Science tech museum	0.083	0.2
Other museum	0.024	0.1
National library	31.751	74.1
Higher education library	1.731	4.0
Public library	0.047	0.1
Special or other library	3.173	7.4
Other organisation	0.032	0.1
All types	42.854	100.0

ANALYSIS OF SURVEY COMMENTS

- 3.87 We have reviewed the comments provided by survey respondents in some detail to establish further context about the statistical analysis. Many either contained explanatory detail relating to answers to particular questions, or some to the opinion about the structure and content of the questionnaire itself. Since we have had to process the response in many different languages, we have categorised the comments made by individual institutions according to the strategic issues they have identified. A fuller analysis of the comments is set out in Appendix IV.
- 3.88 The questionnaire contained extended sections of comments about the approach taken by individual cultural institutions towards digitising their collections: general and technical points concerning the data, use of partnerships, and comments on digitisation contractors. While the general comments provide an insight into the main issues facing institutions, an important element of approach was their attitude towards digitisation. The comments about partners and contractors have allowed us to map the extent and

²² The specific question addressed to institutions was: "How many user requests to access DIGITISED materials were there last year by any other method than the internet? (e.g. CD-Rom within the institution)"

²³ Excludes internet/remote access to the digitised materials.

importance of relationships in progressing digitisation work. The main points we have noted were:

- [i] There is a growing interest in digitisation among institutions holding valuable collections. This tendency is proved by the large number of institutions reporting that although they have not undertaken such work in the past, they are planning to get involved in digitisation initiatives.
- [ii] Digitisation is considered to be a means of raising the visibility and usage of collections. The majority of institutions reported that they were digitising collections to provide improved access. Collaboration patterns – when institutions participated in digitisation networks or coordinated them – also indicated their intentions to make digitised content more accessible to wider user audiences.
- [iii] Institutions were anxious about the constraints on their digitisation aspirations. This was highlighted by the reasons listed for not performing digitisation work, and some negative assessments of current digitisation budgets.
- [iv] There were no significant differences between networks of memory institutions. However, libraries had the most developed partnership networks in comparison to museums and archives.
- [v] There was limited collaboration between archives, libraries and museums. An analysis of the reported networks revealed their general orientation towards collaboration with institutions of the same type. However, archives were more open to collaboration with other memory institutions, while libraries were mostly concentrated on partnerships with institutions of the same type.
- [vi] Institutions usually preferred to outsource digitisation services to commercial enterprises than to partner competence centres. (Some references were made to existing centres in cultural institutions, education and research establishments.)

General comments

3.89 The main topics that arose, and the proportion of respondents raising these issues, were:

- [i] 'state of the art' matters relating to digitisation (56%)
- [ii] the purpose of digitisation (access/preservation) (33%)
- [iii] collaboration (20%)
- [iv] funding (16%).

3.90 A considerable number of institutions were taking their first steps towards digitising collections, and some were pondering further technological developments, or were constrained in proceeding for this reason. Typical descriptions such as 'recently started structured digitisation activities', 'at the beginning of digitising collections', 'we are evaluating what collections to digitise and under what conditions to make them available', 'we will have a plan for digitising in 2009'.

3.91 A typical constraint was the absence of financial resources to support such initiatives. Examples of such comments included: 'due to the lack of funds to purchase necessary technical equipment, so this process [digitisation] has been postponed', 'we have not identified financial resources to start these operations'.

3.92 A minority of institutions expressed their attitude to digitisation in comments about their approach. Of these, a few stated that digitisation was not their priority; therefore they had no plans to digitise or digitisation activities were marginal (e.g. 'digitisation is not our first priority', 'digitisation is not one of the key issues' etc.). Others provided a negative evaluation of their progress, such as, 'not far enough with digitisation'.

3.93 Rather more institutions focused on their digitisation activities to widen access, and rather fewer commented on the preservation issues. However, of those placing priority on widening access, about one-quarter either provided no internet access or restricted access within the institution.

3.94 Institutions usually collaborated in three ways: they contributed content and services into larger projects which ranged from regional and national to international digitisation networks; in some cases, they coordinated such networks themselves and provided technical infrastructure for other content partners; and sometimes, they provided digitisation services to other partners or used external digitisation services themselves. Comments usually included titles of

international arrangements (e.g. Monasterium.net, Europeana, European Film Gateway and others), as well as national projects and networks with other participants.

- 3.95 In raising aspects of their digitisation activities, it was possible to identify the attitude towards current funding arrangements. National and local government were the principal sources of funding. Most institutions reported the arrangements in neutral terms, while a few also provided negative evaluations, e.g. 'present funding is inadequate', 'subsidies... are very limited'.

Digitisation partners

- 3.96 The majority of responses concerning partnership arrangements were made by libraries, although there were also a significant proportion of museums. In particular, the libraries could name more institutions within their networks than other types of institution. References to partners ranged from very abstract (like 'scientific institutions') to specific names. Commercial enterprises (mentioned in about 10% of references) were mostly companies providing information technology solutions; however, in rare cases they represented other cultural sectors not covered in the study framework (e.g. publishing).
- 3.97 Networks and associations made reference to strong professional links. Higher education and research institutions could develop functional associations including the provision of training to staff, advice on specific technology, and specialist collection issues. Some audio-visual institutions appeared to have established partnership links, but in contrast to the libraries and museums, fewer archives were engaged in such partnerships. In rare cases, even schools were cited as project partners.

Digitisation contractors

- 3.98 Not all institutions named their contractors, and several mentioned the commercial sensitivity of doing so. In reviewing these responses, it was evident that a number of individuals were retained as contractors to provide specialist support, e.g. photographers or others performing diverse technical digitisation processes (e.g. image capture, cataloguing etc.). Even so, more than four-fifths of contractors were commercial enterprises. These were mostly digital imaging companies or those providing other IT solutions. It was also difficult to distinguish partnership from contractual arrangements in many cases, particularly when a lead institution acted as the processing centre.

4 REVIEW OF THE METHODOLOGY

4.1 The priorities attaching to the task of establishing a statistical framework differ from those for sustaining and developing a new statistical series. The current study has focused on establishing a network of cooperation between countries. The specificity and consistency that could be adopted from the outset was limited, and only after experiencing some of the lessons of practice is it possible to move on to refining definitions and flexing national priorities to accommodate international standards and procedures for the future.

4.2 Therefore, in reviewing the methodology adopted during this study, we are conscious of the need to outline a way forward, but one which is better informed by the strengths and weaknesses of the approach to establish the base. Based on the experience of the current study there is every prospect of developing:

- * shared understanding of the technical guidance, and the importance of adhering to the sampling and survey administration procedures;
- * a common view about the scope of the survey questionnaire, and the practicability of its implementation;
- * an approach to strengthening the survey response rates.

4.3 Developing recognised procedures and working to common definitions remain the highest priority, and their universal adoption remains paramount. In the meantime, we have reflected on different aspects of the approach in the following paragraphs.

SURVEY INSTRUMENTS

4.4 We employed two main survey instruments. Sheet A has been used to verify the number of institutions in each country, and thereby guide the selection of a representative sample. The other instrument, the survey questionnaire, has been translated into 14 languages, in addition to the base English version.

4.5 On the whole, sheet A has been a very successful instrument in monitoring the survey management standards across all countries. The statistics included in sheet A were essential to place the survey results in proper context. If for no other reason, this instrument has been useful in identifying those countries where there is reliable information about the national cultural domain. It is true that for some countries, the study team have had to research and complete these statistics, but they should now provide a good and convenient

base for each national ministry to review its criteria for defining 'relevancy' and update the base in future. However, as we have previously indicated, consideration needs to be given to whether or not it will be possible to **clarify the definition of the 'relevant' institutions.**

- 4.6 The decision to translate the questionnaire into different languages did extend the preparation time required before mounting the surveys. However, in those countries where no translation was offered, the survey response has been generally lower. It is possible that we would have received a more rapid and higher response if the institutions in these areas perceived that their respective ministries thought the process sufficiently important to merit posing the questions in their national language.
- 4.7 The survey questionnaire was composed of a statistical section (three pages), followed by a request for anecdotal information (albeit structured) about past digitisation projects, partners and specialist contractors. For some respondents, the 'softer' project information sought in the latter part of the questionnaire was welcomed. For others, this part added to the overall appearance of a lengthy and burdensome canvass beyond the capacity of existing institutional information systems.
- 4.8 Therefore, we recognise the benefits of collecting the 'soft' data, but also suggest that if the project information is truly required, it can be collected more efficiently by the interested parties, instead of confusing the statistical exercise for which the information was of limited value. While any **future canvass should be confined to gathering 'hard' data**, the project information gathered in the current survey could be of value to the national ministries concerned where it has not been previously collated.
- 4.9 The design of the questionnaire also incorporated the intention to pose general questions requiring the simplest level of research at the outset (32 questions), followed by a more in-depth analysis of the analogue collections and the progress being made towards digitising these in detail. A shorter questionnaire can be used in future years, to measure the progress made over the current base. The more detailed 'activity' page on the current questionnaire sought invaluable information about the type of materials in collections and how these explain the differences in digitisation costs, yet **it is clear that for many institutions the quantification of their analogue collections remains as problematic as tracing their digital outputs.**
- 4.10 The only domain in which there is an established framework for the international enumeration of the size of institutional collections is in

libraries. While archives have defined measures, there are few available statistics and so few such institutions maintained the relevant data to provide in comparison with the materials they had already digitised.

- 4.11 The value of developing statistics about the size of analogue collections held in museums remains a matter for the European Group on museum statistics to consider. It would also be desirable for any group taking the lead for the collation of statistics covering archive activities to standardise the measures for quantifying the analogue collections.
- 4.12 As a result we believe that ministries will need to rely on making estimates of the progress of digitising the analogue base for some time – unless a more direct incentive can be found to encourage institutions to keep such information. **If the perception in some countries is one where central funding for digitisation work has peaked²⁴, it is clear that a more positive policy incentive needs to be articulated to convince institutions to respond to a survey.**
- 4.13 In the meantime, the detailed breakdown of the analogue and digital collections provided in response to the current survey was sufficient for current research purposes. We do not believe that it will serve the sustainability of headline statistics to collect such detailed information within the framework in the foreseeable future.
- 4.14 Gaining responses from individual institutions was always seen as a challenge and, in anticipation of this, the support of the national coordinators was essential. A cross-domain base has been established, but it is recognised that this has required considerable effort on behalf of both those institutions supplying the data, and the national coordinators encouraging them to do so. Sustaining the statistical framework will depend upon simplification, standardisation and the award of higher status to the exercise.
- 4.15 The **importance accorded to the statistical framework is crucial** if it is to be properly implemented at the national level. This is a problematic point, given that the only mandate on the ministries in the respective community countries was a shared vision of the importance of establishing the statistical series. The priority devoted to that vision at ministerial level can only be encouraged with the passage of time and an insight into the benefits that a series, based on the publication of the baseline results, promises to provide.

²⁴ According to anecdotal evidence there is the feeling in some institutions in a number of countries that the specific funding for digitisation has peaked. Some ministries may find alternative incentives to respond to surveys, perhaps by identifying progressive institutions as exemplars.

- 4.16 Not all the relevant cultural ministries adopted the prominent role in managing the framework; in 14 of the 27 member states, the cultural ministry engaged leaders in the professional networks or agencies to encourage their institutions to participate in the surveys. Furthermore, the simple realities of administration meant that the support needed to chase the survey response and scrutinise the data were additional activities within previously fixed resources. While several of the cultural ministries had well developed internal statistical frameworks, the more direct involvement of the central statistical division or agencies in many others would benefit the arrangements.
- 4.17 Future arrangements depend on direct ministerial acknowledgement and support for the framework in all countries, even if the management of the framework is undertaken by another agency.
- 4.18 To take forward the current approach, a number of countries found it necessary to adapt the guidelines to accommodate their own circumstances. Some undertook the data collection phase themselves, but restricted the target sample to focus on the main institutions rather than the broader guideline sample.
- 4.19 In the short term, it is unlikely that all countries will modify their existing statutory returns to accommodate the required survey data given the level of detail required. Thus, although **a more efficient manner for collecting data in the longer-term future may be via such national cultural policy / planning returns, this approach may lead to delays in pooling international data**, and may weaken the adoption of consistent definitions, as the policies of each country will inevitably differ.
- 4.20 Similarly, any modification to the framework questionnaire leads to different definitions – either alternative groupings of institutions, or data definitions; for example, the collection of category rather than scale data (number of institutions whose data fall within broad ranges, rather than absolute values). This clearly hinders the calculation of derived ratios and the scope of further analysis. Indeed, where the administrative responsibility for particular ministries does not cover the whole cultural domain, then it is clear that alternative arrangements are necessary to ensure that all institutional digitisation activity is surveyed.
- 4.21 Simplicity is clearly of the essence. Therefore, the approach for future years needs to be sufficiently uncomplicated to be capable: (1) of endorsement by all ministries; and (2) of being presented in a manner which capitalises on the fact that the initial effort to establish the base provides for a less intensive approach to sustain the series. We believe that this can be achieved if the responsible

ministries agree to collect information from their 'relevant' institutions (using **a shorter questionnaire** either addressed to all, or a sample), about the volume and cost of digitisation undertaken in the previous year.

- 4.22 To ensure the success of this approach, **a coordinator is required at the European level. The European Commission has a leading role for coordinating the framework, but the ministries in each country would need to take direct responsibility for the canvass** in their own countries. The 'simplicity' of the canvass for data and the standardisation of relevant institutions from which the information is sought is discussed below.

DEFINITIONS

- 4.23 The study methodology stressed the need to make a random sample of institutions to ensure that a representative picture could emerge of the relative progress being made within the universe of identified relevant institutions. Some countries followed the guidelines; others did not. This places the primary **focus on developing a shared understanding of which institutions to treat as relevant** in the future, to enable the more forensic approach to sampling to be implemented properly. In the absence of any better definition, the study methodology necessarily focused on a practicable definition of the relevant universe of institutions capable of generating significant value towards the nation's digital heritage, but future guidelines would benefit from further definition.
- 4.24 A recognised definition of 'digitisation' had been developed by the Institute of Museum and Library Services²⁵, and we adopted this in our approach: "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". While this definition is concise, ambiguity exists over where the start and finish lie. To avoid misunderstanding, we extended the definitions in the questionnaire to clarify how to treat the management and supervision of digitisation projects, and the extension of the definition to finishing tasks, such as the interpretation of digital images and formatting these for the web.
- 4.25 However, if the institutions involved are to more consistently report their staff time spent on digitising materials, and the costs involved, more practical guidance is needed to distinguish the many different processes involved in the journey between transferring materials

²⁵ IMLS, *Status of Technology and Digitization in the Nation's Museums and Libraries*, 2006, available at: <http://www.imls.gov/resources/TechDig05/Technology%2BDigitization.pdf>

onto electronic media, and then transforming the digital image into an interpretable format. A **checklist of the digitisation common processes would considerably assist institutions in arriving at more consistent judgements of the time and cost involved** for their own planning purposes. This, in turn, will lead to the adoption of more consistent interpretations of the statistical definitions, i.e. better statistical definitions, and better adherence to these will only arise if the definition coincides with the management information requirement.

- 4.26 The specific definitions we adopted for questions applying to the resourcing of digitisation had been developed after extensive testing. On the whole, the financial questions were well completed (see completion rates described in section 3). However, while it would be possible to refine these further, it should be born in mind that unless the surveys are completed by staff familiar with the resources management issues within their own institutions, such refinement would simply add further complexity. Apart from those areas listed under *Other priorities* below, the remaining definitions were fit for purpose, and give a good base for any further framework to develop. They were:

Institutional staff

Equivalent full-time staff employed by the institution, assuming part-time staff were converted to the equivalent number for a full working week. This included both permanent and temporary staff on the establishment, but excluded contractors' staff.

Annual budget of the institution

This was the approved revenue budget, including the annual provision for staff and running costs of the whole institution, including the cost of planned purchases to replace items in the collections, depreciation, maintenance of IT etc.

Digitisation staff

The term included the time of the institution's own staff engaged on digitisation projects, such as: planning and managing in-house and contracted projects; preparing and digitising materials; and enhancing digitised output to widen accessibility.

Digitisation budget

This sum was the amount identified in the institution's own annual budget to provide for digitisation work. Respondents were asked to include the cost of the staff time devoted to digitisation projects in

this estimate, but to exclude external funding, if this was not accounted for by the institution.

Government funding

Grants made by central, regional or municipal government for specific purposes, of which digitisation was an approved category.

In-kind support

This included other funding not categorised elsewhere, including that deriving from commercial arrangements.

Cost of planned digitisation

This was the equivalent of either the institution's estimate of what a contractor would charge to digitise the materials, or the financial resources identified in the plans. Where a plan existed, this was defined as a written strategy document covering the digitisation of materials held by the institution for any period up to 2012.

- 4.27 The definitions provided for the measures of digitisation progress were also sufficiently practicable for a good proportion of survey respondents to provide their estimates.

Part of the collection where there is no need to digitise the materials

The part of the institution's collection that did not form an important part of the nation's cultural heritage, or would be duplicated²⁶ – has been (or will be) digitised by another institution. (Digitisation of reference items obviated the need to digitise everything.)

Part of the collection that has already been digitised

This included projects that were approaching completion. Materials that had been recorded as images, even without character recognition, were to be included as digitised materials.

Part of the collection that still needed to be digitised

This was the outstanding part identified as requiring digitisation, either to assist with preservation, and/or because the digitised material would sufficiently improve open accessibility to a larger clientele.

²⁶ Note that the definition attempts to eliminate the digitisation of materials commonly held in many institutions, so that duplication is minimised in the deriving statistics.

OTHER PRIORITIES

- 4.28 Within the detail of the questionnaire, there were, however, five other particular areas where the definitions need particular development. These relate to:
- * newspapers;
 - * monuments;
 - * digitisation in museums (catalogue/materials);
 - * records in archives;
 - * user access to digitised materials.
- 4.29 The issue with newspapers is problematic. We tried several different collective measures for newspapers when testing the questionnaire. None were wholly satisfactory. Institutions tend to measure their newspapers in terms of 'collections', 'editions', or 'issues'. **Gaining a consistent count depends upon further consideration by the standing ISO group and/or a group of interested national coordinators.**
- 4.30 Including monuments in the more detailed survey analysis of materials that are being digitised was not a complete success. The definitions exist, but the diversity in types of monument inevitably leads to alternative views on the definition of digitisation for each type. Furthermore such monuments can either be part of a wider heritage site, falling under local administrative control, whilst others are maintained by a central agency. The approach to measuring the digitisation of these objects needs to be reviewed.
- 4.31 We have made reference to the doubts expressed by several observers that the measure of digitisation has been interpreted consistently amongst museums. The definition may need reviewing to ensure that responses to future surveys measure the digitisation of analogue objects held in museums in the same ways as the assessment is made for collections held in other types, such as libraries and archives.
- 4.32 Another type of difficulty attaches to the accurate recording of the digitisation of materials held in archives. The study has revealed just how differently the analogue collections are recorded in this type of institution. The particular difficulty remains to gain consistency in this area, so that future surveys can relate the digital output to a consistent measure of the analogue base.

4.33 User access is perhaps the most difficult measure to gain. The definitions exist, but these need to be improved to accommodate the paucity of adoption of common standards in institutional management information systems.

STRUCTURAL ISSUES

4.34 The cultural heritage may include the monuments of the nation, but few countries sought such information. In our view, most digitisation work relating to monuments falls under the heading of 'cataloguing' or 'born digital', which both fall outside the scope of the brief about which the statistical series is intended.

4.35 Following the workshop in April when the scope of the survey was agreed, we produced a separate questionnaire that each national ministry could employ to collect its own supplementary data from agencies responsible for the upkeep of monuments. Seven countries²⁷ have used this specific questionnaire, but the response has been insignificant. In our view, the development of the framework should focus on the materials that would add value to the extension of 'accessible information' as a priority. The facility to collect such information in future remains in the special questionnaire designed for the purpose.

4.36 The problem with the reporting of archive collections in a consistent manner has been described previously. Since the difficulty arises from the choice of unit for measuring the size of collections (shelf length, volumes, collections, boxes), it may be simpler in future to seek data by the number of pages alone. User access definitions exist, but they are not yet applied consistently.

DIVERSITY WITHIN THE CULTURAL SECTOR

4.37 A significant issue to emerge from the current data collection exercise is the real difference between types of cultural institution in their interpretation of both the meaning of and the scope of their digitisation processes. The current high level definition of 'digitisation' does not disguise the fact that different standards exist for different types of material. For example, even within one institution, the dilemma facing the respondent was one of defining the standard for the measure: either '13% of the archive is now digitally preserved to minimum standards', or '1% of the archive is digitally preserved according to best international practice'.

²⁷ Two institutions in the Czech Republic; one in Finland; three in Greece; two in Romania; three in Slovakia; one in Slovenia; one in Sweden.

- 4.38 The definition of 'digitisation' is key to the funding debate. Agreeing definitions across all domains at international level is a major task, but it is crucial to the availability of statistics in which institutional stakeholders can recognise the value. More, and urgent, work is required in this area requiring the consideration an expert group, and we also believe that useful guidance to individual institutions in the form of **a costing checklist to conveniently summarise the relevant processes should be included in the definition of 'digitisation'**.
- 4.39 Regardless, **the current approach using a generic questionnaire has the advantage of seeking consistency across all types of institution.** However, the simple average describes a very broad range of values across all institutional types. Furthermore, the domains are quite distinct in many respects; it is not possible to add the number of hours of audio-visual output to the number of pages of digitised text and the number of images of objects, and gain a meaningful summary measure for all cultural institutions. This emphasises the need to develop the statistics in such a way as to measure the relative progress in the different domains.

5 PROPOSALS FOR SUSTAINING THE FRAMEWORK

- 5.1 It is our belief that simplicity is required, both in the approach to collecting data in future years, and in the volume of data collected. The majority view of national coordinators providing us with their feedback endorses the need for brevity, but for several, there is an appetite for more detailed statistics. The relative priorities described by these national coordinators was as follows.

PRIORITIES EXPRESSED BY NATIONAL COORDINATORS

- 5.2 **AUSTRIA:** Simplify the questionnaire, and rationalise the amount and specificity of the financial information being sought. The overall response to the current survey may have been improved if it had been addressed to the director of the institution, rather than the professional responsible for the area.
- 5.3 **BELGIUM:** Clarify the definition of 'relevant' institutions, including recognisable criteria for their selection in order to better ensure a consistent universe within which the progress of digitisation can be monitored.
- 5.4 **BULGARIA:** Place more emphasis on qualitative rather than quantitative parameters. For instance, seek indicative financial estimations, rather than precise costings for which detailed definitions would be required. Any demonstration of the links between the survey and the funding of digitisation work would benefit the response rate. The European Commission's support of an ongoing network developing definitions and the survey standards would help to demonstrate this. A particular difficulty encountered during the current study was the disparate nature of many organisations. This made it difficult to identify and target the survey questionnaire effectively. Providing the headline approach was supported centrally, individual ministries and national coordinators would be able to devote their time on the greater detail they required.
- 5.5 **DENMARK:** Seek fewer details; set out better definitions; develop a more focused survey.
- 5.6 **FRANCE:** Future surveys might focus on the progress being made towards the digitisation of priority value areas rather than the complete range of materials held in collections. Consistency in the selection criteria adopted for sampling will be essential. In the meantime, comparative measures of the access made to digitised collections were insufficiently reliable to publish.

- 5.7 The framework would benefit from the involvement of Eurostat, and a shorter questionnaire would be preferred.
- 5.8 GREECE: Future questionnaires should be more abstractive and brief, in order to lessen the burden on institutions and boost the rate of response.
- 5.9 IRELAND: A simplified version of the questionnaire, removing all but strictly necessary questions, would improve the response rate. However, some doubt remains as to whether or not an annual survey is required.
- 5.10 LITHUANIA: The difficulty of gaining responses from individual institutions was emphasised, pointing to the preference for the use of a less daunting survey questionnaire in any future framework.
- 5.11 NETHERLANDS: There was a view that the current survey was too broad to provide a recognisable focus, but instead of reducing the number of questions asked, it was considered more appropriate to mount a number of different surveys on each topic. The feeling was that a shorter questionnaire would benefit the policy makers, but would be insufficient to provide individual institutions with the appropriate benchmarks that interested them. The breadth of the current questionnaire was seen as a good incentive for the institutions to invest in (better) management information systems of their digitisation efforts.
- 5.12 SLOVENIA: It was suggested that institutions would benefit from receiving any questionnaire a year in advance of the response date, to provide sufficient time to marshal the required information and carry out the necessary research.
- 5.13 We do not believe that there is a real dichotomy in the above views, and we do think that they can be easily reconciled. The desire for more detail could be accommodated, by **concentrating the development of the international framework on headline statistics; and those nations intent on researching progress in more detail could encourage specific benchmarking initiatives** (see recommendation E in section 6).

STREAMLINING THE FUTURE SURVEY REQUIREMENT

- 5.14 A refined and shorter standard questionnaire should be agreed (see recommendation C). The manner in which the information is then collected from individual institutions should be a matter for each

ministry to decide, providing the data can be made available to the internationally agreed timetable and to the same definitions as the standard questionnaire.

5.15 The information should be collected from all those institutions that responded to the current survey, so that the movement from that baseline could be more easily assessed. A selection of more institutions in those countries where the response to the current survey was sparse would be made in order to improve the reliability of the baseline and the ongoing series of statistics.

5.16 The refined questionnaire might request data on the following three 'essentials': what resources *inputs* are or will be required; what *outputs* have been or are planned to be achieved; and what *outcomes* are arising in terms of improved user access:

- * total expenditure on digitising materials in the previous 12 months; total expenditure expected to be made in the next 12 months on digitising materials (or 24/36 months, if the frequency of the survey for updating the series is less than annual);
- * volume of digitised output last year – pages, objects, collections, hours, as appropriate; volume of planned digitised output in the next 12 months on digitising materials (or 24/36 months, if the frequency of the survey for updating the series is less than annual);
- * number of online user visits to access digitised materials in the previous 12 months; or a similar measure that can be defined in a manner which attracts more consistent interpretation; and, possibly
- * income arising from the sale of access to digitised materials, during the previous 12 months.

5.17 This final requirement may not be strictly necessary, but we have noted from our direct communication with many institutions in connection with the current survey that there is a high degree of cynicism regarding the benefits that their contribution of data can make to gaining resources for their activities. With the above priorities in mind, we set out below an indication of the form a shortened framework survey might take:

SHORT 'CORE' QUESTIONNAIRE

Investment in digitisation

1 Total expenditure on digitisation: (and the period to which it relates)

[i] Last year: > Period:
 [ii] Current year: > Period:

2 Full-time equivalent staff engaged on digitisation projects:

Current year: > Period:

3 Cost of planned digitisation projects:

Project spend: > Period:

Outputs

4 Percentage of the collection that has already been digitised: %

5 Volume of analogue material digitised in the last year:

[i] Pages: of written/image material
 [ii] Hours: of audio-visual materials
 [iii] Metres: of archive material not included at [i]
 [iv] Number: of all other objects in collections

6 Volume of analogue material to be digitised in plans described at question 3:

[i] Pages: of written/image material
 [ii] Hours: of audio-visual materials
 [iii] Metres: of archive material not included at [i]
 [iv] Number: of all other objects in collections

Access to digitised material

7 Number of user visits to access digitised materials last year:

[i] 'Free' online:
 [ii] Off-line (any):

5.18 The example is a draft for further consideration, and particularly in the light of the difficulty we have experienced in gaining reliable responses to the measure of user visits/accesses to digitised materials, the final form could be quite different. However, there are several advantages to such a form: its brevity presents the

survey objectives clearly, and as a consequence it is less daunting than the previous questionnaire. Furthermore, the questionnaire employed in the study just completed was sufficiently extensive and broad ranging that it required extensive research within the organisation, often requiring the input from persons engaged in different departments of the same institution; inevitably leading to delays and mistakes in its completion. **A shorter survey such as this could be more easily administered online**, and so extended to a larger sample, or all relevant institutions.

FOCUSING THE FUTURE SURVEY REQUIREMENT

5.19 **The criteria for identifying those institutions that are relevant to the measurement of digitisation effort need to be agreed** at an international level to ensure the consistency and uniformity of future surveys. We make later recommendations for gaining this agreement, but suggest the following points for such consideration:

- * "Relevancy" relates to institutions holding collections of cultural importance to the nation whether or not a start has been made on their digitisation. Cultural materials of national importance can, of course, form part of the collections of provincial institutions; and will not necessarily be found exclusively in the "National" institutions.
- * These "Relevant" institutions will in the main be Archives, Museums and Libraries, but in so far as their collections contain material of value to a national digital collection, will also include film institutes, broadcasters and any other agency or organisation with responsibility to maintain such collections.
- * The terms, "Library", "Museum" and "Archive" may be distinct definitions of general purpose, but particularly at national level, many institutions span some or all of the services normally dispensed by each. So to avoid splitting hairs, we suggest that multi-purpose organisations are defined according to their prominent collections. This was the definition we adopted in the current study, but we observed a tendency amongst responders to "split hairs". Whatever criteria are adopted, it is clear that national coordinators will have to take responsibility for imposing the agreed definition in future. For consistency in the classification of the collected statistics, there needs to be uniformity in the way institutions are classified.

- * Organisations providing a centralised digitisation service will not be “relevant” unless they hold significant collections in their own right. However, these agencies will need to be identified and surveyed separately to ensure that the volume of digitisation work they undertake for the “relevant institutions” is correctly assessed.
- * The “Relevant Archives” are mainly those that hold records of national, provincial or municipal government. The established definitions (ISO 5127 and Pierce-Moses)²⁸ are too wide in their scope, referring to all organisations with departments that maintain records of “enduring value”. Instead, the selection of archives should focus on the records of state. (Film and other types of audio-visual “Archives” are considered separately below).
- * The “Relevant Museums” are not exclusively state institutions, as many national cultural treasures are housed in collections maintained in trust. Never-the-less a distinction needs to be made between the many local institutions and those that provide a significant provincial locus for national collections. The national and main provincial museums should be included. The “Main” provincial institutions would be those serving a regional area, rather than a municipality, unless that municipal institution served a wider visitor interest or specialist national collection.
- * The “Relevant Libraries” would be those serving a national interest in that a significant part of their collections are the sole sources of national texts. The ISO definitions make reference to different types of library. We suggest within these definitions the following are excluded: Special Libraries, unless their collections are unique. Public Libraries, unless these are the main service point for a region within the country. In some countries the Region could be defined by population size, but in many smaller countries the geography is more important determinant. All School libraries should be excluded; and only those libraries of higher education institutions should included if their collections are unique and form a major part of the nation’s cultural heritage.
- * Audio-visual institutions are important organisations that hold “Archives”. As observed above, this causes misunderstanding as far as the grouping of organisations is concerned. However, for the purpose of identifying institutions that contribute significantly to the cultural heritage of the nation,

²⁸ See appendix III, annexe 1 for these definitions.

the national ministry will need to consider whether the collections held by these institutions actually hold the value (to the national digital heritage). The difficulty encountered in our study, was that some countries believed they did, some did but did not include them, and some did not and excluded them. This problem will persist unless all adopt a simple approach, as follows: All institutions, whether describing themselves as "Archives" or not, that hold audio-visual materials in their collections that are considered of national importance (by the Ministry responsible for of "Culture") should be included as "Relevant". As a minimum, we suggest that all members of the Association des Cinémathèques Européennes (ACE), or the Fédération Internationale des Archives du Film (FIAF), should be included as relevant institutions.

- * We suggest that the national organisation appointed to be responsible for the broadcasting of news and information in the nation (and recognised as so by the Ministry for Culture) should be included.
- * There are a number of other institutions responsible for the maintenance of "Cultural assets" that may be considered as "Relevant". We recognise that there may be ministerial agencies responsible for monuments. The response from such agencies / institutions to the survey conducted in the current study was insignificant, even amongst those countries that had strongly argued for their inclusion. We recognise the importance of monuments in the cultural domain, but believe their inclusion in a canvass of digitisation activity to be a distraction. However, we understand that such an area (and others) will be of particular interest and value to different nations. Therefore, we suggest that a category of "Relevant" institution know as "Other" be included, as provided for in the current study. However, we suggest that each national coordinator provides a proper justification of how the collection administered or maintained by the agency makes / could make a significant contribution to the digital heritage of the nation, before such institutions are included in any count of "Relevant institutions".
- * We recognise that in many institutions, the financial and other strategic management information will be held by other administering offices. For this reason, some national coordinators in the current survey just completed directed questionnaires to the administering municipality rather than the institution. This distracted from the proper identification of the number of relevant institutions, and care should be

taken in future to establish to which institutions any data collected in this manner relates.

5.20 We understand the difficulty of adopting too complex rules, as to make the definition impracticable. A particular problem arises for the individual attempting to identify such institutions, if he or she has limited knowledge about the institutions listed in any national register; even if such lists exist. Therefore, to summarise these points the following framework may serve as a starting point for refinement: The "Relevant" institutions are:

- (a) Archives that maintain records (act as documentation centres) of central or local government of particular heritage value. i.e. the national, provincial and significant municipal records offices. Administrations serving populations of less than 5% of the national population would normally be excluded.
- (b) All nationally prominent Museums and "Popular cultural centres" including provincial institutions serving a regional area, but excluding municipal collections, unless that municipal institution either served a wider visitor interest equivalent to 5% of the national population, or housed a unique and specialist national collection.
- (c) Libraries would include at least all member institutions of the Conference of European National Librarians (CENL), and other recognised "national" libraries in the country. Only those libraries housing unique national collections should be included amongst those attached to Higher Education institutions. Libraries attached to or serving professional bodies or significant special interests whose collections are of particular national and historical interest should be included. Most "Popular" or "Public" libraries would be excluded unless the institution provided a reference / lending function to more than 5% of the national population. All other libraries would be excluded.
- (d) Public broadcasting institutions, but excluding all other broadcasters. All members of the Association des Cinémathèques Européennes (ACE), or the Fédération Internationale des Archives du Film (FIAF), should also be included as relevant institutions.
- (e) Some other institutions, providing they were responsible for maintaining or housing significant cultural assets of national value should be included. In this sense, any ministry or agency should only be included if the care for

these assets was not devolved to another institution within their administration.

- 5.21 The **database of relevant institutions (as represented in sheet A) should be updated** by national coordinators to take account of intervening changes (see recommendation C). This should be an **annual exercise** and, in its own right, should be capable of providing **some practical measures for Eurostat to describe the cultural sector in Europe**. The list of contact details could also be usefully extended to all (rather than a sample) of relevant institutions so that a single international agent could undertake surveys directly on behalf of the national ministries.
- 5.22 In the meantime, a specific **group of experts should be asked to consider whether it would be possible to develop a more specific definition of relevant institutions** (using the suggested criteria set out above), that would lead to greater international consistency (see recommendation B). It is appreciated that there are a number of standing groups representing the international communities of particular cultural domains, such as the European Group on Museums Statistics, but the specific focus of the Numeric framework transcends these narrower interests. The framework would benefit from an identified group to take ownership of developing standards and definitions for the future framework.
- 5.23 **A commitment should be gained from each ministry** (either that responsible for national statistics, or for cultural affairs) to support the framework, **in a manner that endorses the position of national coordinators** in encouraging institutions to respond in the national (rather than simply the professional) interest. Their endorsement of a special interest group, referred to in recommendation A, is implicit in this proposal.
- 5.24 In this connection the interests of both individual ministries and the suggested special interest group would be better served if the European Commission provided **a facilitator to coordinate the survey activities in future**. We suggest that such a facilitator would work to broad terms of reference as follows:
- (a) Service a special interest group of experts (as suggested at 5.22) to advance the framework, refine definitions and agree future survey instruments;
 - (b) Maintain a register of national coordinators in each country and provide advice to each regarding the implementation of the framework, including specific guidance about the identification of relevant institutions. Encourage national coordinators to survey all relevant institutions rather than

spend further time in designing a representative sample, if this approach is agreed by special interest group.

- (c) Draw up the outline survey instruments and supporting tools, and once approved by the special interest group, then distribute these to national coordinators and agree with them timetables for the undertaking of the survey in each country;
- (d) Alert the special interest group to any problematic issues and provide advice for resolving such difficulties;
- (e) Provide specifications to each country for mounting on-line surveys hosted on the national coordinators' web-site of choice;
- (f) Undertake any extended survey on behalf of international groups of professional interests (e.g. CENL) and liaise with each national coordinator accordingly to ensure that the collection of such data is harmonised with their own undertaking of specific national surveys;
- (g) Review specific proposals made by any single country to develop more in-depth survey approaches, including benchmarking initiatives, to provide effective clearing-house arrangements for any joint survey design / processing arrangements;
- (h) Provide a central survey data processing capability, for those countries opting to refer their institutions to return data to the central facilitator. Act as the manager of the central database for data collected directly in each country;
- (i) Review all data collected, and refer queries (plausibility and other cross referencing validation checks) to national coordinators for resolution;
- (j) Analyse the pooled data arising from all surveys on behalf of the special interest group.

5.25 Providing the timetable for agreeing the survey instruments and implementing the surveys is concluded to a common timetable, these tasks could be pursued in an efficient manner. However, in our experience this coordination role is likely to require the availability of a facilitator throughout the whole year, since national priorities and programmes will vary. Inevitably, the probability of 27 countries working in exact step remains remote, and so the workload should not be underestimated. Furthermore, the

facilitation role is seen as a technical support role, and would need to be effectively supported by the European Commission, given the different administrative protocols each national coordinator will necessarily need to follow. The scope of the facilitator's responsibilities could extend to providing bespoke survey instruments incorporating appropriate language translations and then writing a full report on the results; but this would expand the role well beyond the technical support role we envisage as necessary to develop the international statistical framework.

- 5.26 A review should be undertaken, six months after the completion of the next survey round, to **ascertain the use made of the survey results in each country**. In the meantime, the follow-up survey should seek to collect updated information from those institutions that participated in the current survey, to better measure the trends. This survey should also be extended to those institutions in the recast 'relevant' base that either did not respond before, or were not included in the previous sample.
- 5.27 Lastly, as the current study has progressed we have been struck by the potential for misunderstanding in the interpretation of the technical issues concerning statistics. By way of an example, we have used the 'median' measure to summarise the distribution of data describing the diversity of institutions. Some have interpreted this as a failure to direct the sample at institutions with more homogeneous characteristics, i.e. with a small variation about the arithmetic mean. Indeed, we have encountered at least one justification for drawing national samples from only the main institutions, so that the statistics describe a less dispersed universe. This rather defeats the object, since scientific method demands the measurement of the real cultural sphere, where one of its attractions is its diversity.
- 5.28 In this connection, our note of caution is simply that any single measure is not capable of describing this variety. Our guidelines for selecting a sample that was meant to be representative of this diversity were drafted for good reasons, and it is hoped that the principles we had in mind will be adhered to in future.

6 RECOMMENDATIONS AND IMPLICATIONS

- 6.1 The rationale for the following recommendations has been argued throughout the report, and summarised in the previous section. We are, however, conscious that the implications of any action plan to develop the framework should be properly understood, to ensure that such a plan is sustained by adequate support from the stakeholders concerned.
- 6.2 We have several recommendations relating to how the process is supported in the future, and others, of a technical nature, concerned with the design of the framework. Each recommendation is denoted by a prefix letter [in square brackets].
- 6.3 **[A] To endorse the principle of a special interest group dedicated to developing the standards and definitions for future survey activities** (a standing technical group on digitisation statistics). Whilst this group could meet virtually, it is important that those involved can represent the views of their respective national ministries. Our views on providing support for this group and the coordination of the framework are outlined at 5.24 above.
- 6.4 There are several existing groups representing the interests of particular domains, and although these have the required expertise, it is particularly important to establish a forum which can develop the approach for all domains in concert. Furthermore, the support of agencies such as Eurostat, and the broader interest of the Institute of Statistics at UNESCO, can more efficiently engage with a cross-domain group. Given the technical nature of their agenda, a core membership of the special interest group should be considered to represent the interests of all countries.
- 6.5 **[B] An agenda for the special interest group should be specified prior to its formation**, so that it is initially constituted to effectively address the priority issues. We recommend the following focus on definitions and standards:
- * Define the criteria for identifying 'relevant' institutions. Given the desirability of measuring progress, it is important that those institutions with collections that have not yet begun a programme of digitisation are not excluded, providing of course their materials are of significant value to the nation's cultural heritage. Our suggestions for considering the development of the definition are set out at 5.19 above.
 - * Review the most appropriate practicable means of measuring access to digitised materials.

- * Consider more suitable definitions for the other most problematic measures encountered during the study (the units for newspapers and records in archives), and inform the appropriate ISO technical group/international professional organisation accordingly.
- * Develop the definition of 'digitisation' to include the relevant checklist of processes to be included under the heading. (This has wider implications for the management information systems of all institutions.)

6.6 **[C] Provide national coordinators with a draft specification, including outline survey instruments for a short survey scheme**, in order to consult on the timetable and form for updating and developing the statistical base. This consultation needs to proceed apace with the deliberations of the special interest group on digitisation statistics. The essential instruments include:

- * short questionnaire;
- * guidance on the identification of relevant institutions;
- * specific procedures for the consistent selection of impartial and representative sample institutions to include in the survey – incorporating revisions for sheet A, the vehicle for recording relevant institutions, and the sample;
- * arrangements for administering the survey online;
- * checklists for the encouraging and chasing survey response.

The agreed instruments should then be described in **a newsletter** that can be circulated to **relevant cultural institutions** in each country, so that they can **prepare the required management information in advance** of any coordinated international canvass or 'developed' statutory return. The practicability of administering a further survey during 2009, to provide results that can be reported in 2010, to coincide with the reporting arrangements of the Member States' Experts Group, needs to be considered in this context.

6.7 **[D] Provide the facility for the principal professional associations** (e.g. Conference of European National Librarians) to employ the framework survey to collect additional data that they would otherwise need to collect themselves. In such instances variations to a core questionnaire would be restricted to the type of institution represented by the specific professional group whose data needs had been incorporated by the special interest group.

- 6.8 **[E] Advertise the potential interest that a number of national agencies have expressed in developing benchmarking initiatives** among the institutions in their country. It is likely that each of these will have distinct objectives, but there are equally likely to be common elements, providing the opportunity to design a summary specification that generates the headline data required for the Numeric framework.



ACKNOWLEDGEMENT:

The study team would like to express their thanks to all national coordinators for 'volunteering' their efforts towards establishing the statistical base represented by the current study. In this connection, we wish to inform the MSEG of the valuable contribution of their nominees towards establishing the statistical series. (Please refer to appendix V.)

There were also many other persons in each country, too numerous to list, who provided good advice and assistance to the shaping and undertaking of the survey. Not least, we wish to thank all those persons who completed the questionnaire with such patience and diligence. Their efforts bear witness to the remarkable international goodwill that exists amongst those persons dedicated to the management of the cultural institutions in Europe.



KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

Access, 3, 6, 9, 14, 17, 18, 19, 28, 42, 60, 61, 62, 63, 64, 68, 70, 71, 72, 73, 85, 87, 88, 96

AIFF, 58

All types, 25, 26, 28, 33, 36, 37, 42, 62, 84

Archives, 6, 7, 8, 14, 21, 25, 26, 28, 29, 34, 39, 45, 47, 49, 52, 62, 63, 64, 69, 70, 72, 74, 77, 82, 97

Art museum, 59

Artefacts, 17, 65

Audio-visual, 3, 7, 8, 16, 25, 27, 34, 49, 51, 56, 59, 62, 66, 69, 70, 71, 74, 84, 88

Austria, 21

AVI, 58

Backlog, 51, 52

Belgium, 21, 31

Benchmarking, 7, 10, 86, 98

BETA, 57

BMP, 57

Books, 8, 17, 65, 79

Broadcaster, 27

Budget, 3, 38, 40, 43, 44, 46, 48, 80

Bulgaria, 21

CD, 18, 57, 63

Central Statistical Offices, 20

Chiffres Clés, 25

Collaboration, 72, 73

Commercial, 49, 72, 74, 81

Conference of European National Librarians (CENL), 7, 16

Consultations, 17

Contractors, 45, 49, 51, 69, 71, 74, 76, 80

Cyprus, 21

Definitions, 6, 9, 11, 13, 15, 20, 75, 78, 79, 80, 81, 82, 83, 84, 85,

KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

87, 93, 96, 97

Denmark, 21

Digitaal Erfgoed Nederland (DEN), 27

Digitisation, 1, 2, 3, 6, 7, 8, 9, 10, 11, 13, 14, 15, 18, 19, 21, 22, 24, 25, 26, 27, 28, 31, 36, 37, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 52, 54, 57, 58, 59, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 88, 96, 97

Distribution, 3, 20, 23, 30, 33, 34, 35, 37, 44, 46, 49, 61, 64, 66, 67, 68, 71, 95

DjVU, 57

Domains, 7, 13, 22, 25, 34, 38, 62, 84, 93, 96

Drawings, 17, 65

DV, 57

DVD, 57

e-Mail, 20

Engravings, 17, 65

Error, 6, 11, 33, 38

Estonia, 21

Europeana, 18, 74

Eurostat, 86, 93, 96

Excel, 57

Exchange rates, 39, 44, 69

Federal, 21, 23, 24

Film archive, 9, 28, 70

Film institute, 6, 49, 51, 52, 59, 60, 61, 64, 66, 69, 71

Finland, 21, 83

FLAC, 57

Flemish, 24

Format, 57, 58, 80

France, 21, 31

French, 24, 25, 39

Full-time equivalent, 3, 45, 46, 88

KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

Funding, 3, 1, 10, 11, 15, 24, 26, 40, 44, 49, 50, 73, 74, 77, 81, 84, 85

German, 24, 39

Germany, 21, 33

Government publications, 17

Grande Duchy (of Luxembourg), 27

Greece, 21, 83

Guidance, 2, 14, 19, 20, 39, 75, 79, 84, 97

Higher education library, 66

HML, 57

HTML, 57

Hungary, 21

Images, 3, 9, 16, 17, 55, 57, 58, 59, 79, 81, 84

In-kind support, 49, 81

Input, 9, 16, 68

International comparisons, 11

International Standards Organisation (ISO), 9

Internet, 3, 8, 18, 60, 62, 63, 70, 71, 73

Ireland, 21

IT resources, 15

Italy, 21

Journals, 8, 17, 79

JPEG, 57, 58

Latvia, 21

Libraries, 6, 7, 8, 9, 14, 21, 25, 26, 28, 29, 30, 45, 46, 47, 49, 51, 52, 57, 59, 60, 61, 62, 63, 64, 66, 70, 72, 74, 77, 79

Lithuania, 21

Long-term preservation, 47

Luxembourg (see Grand Duchy), 13, 18, 21, 27

Malta, 13, 21, 27

Management information, 6, 8, 9, 15, 17, 22, 39, 80, 86, 97

KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

Manuscripts, 17, 65

Maps, 17, 65

Master format (see "Format"), 57

Metres, 16, 17, 54, 65, 88

Michael / Minerva, 25, 26

Ministère de la Culture et de la Communication, 21, 25

Ministry, 15, 21, 25, 26, 27, 28, 66, 76, 78, 83, 87, 93

Monuments, 8, 17, 18, 26, 33, 65, 66, 67, 82, 83

Motion, 3, 58

MP, 57, 58

MPEG, 57, 58

Municipal, 28, 49, 81

Museums, 6, 7, 8, 14, 17, 21, 25, 26, 28, 30, 39, 47, 52, 59, 62, 64, 70, 72, 74, 77, 79, 82, 93

Music, 17, 65

MXF, 57, 58

National Library, 21, 28, 39, 66

NEF, 57

Netherlands, 21, 27

Newspapers, 8, 17, 65, 82, 97

Number, 14, 16, 17, 18, 19, 22, 23, 24, 27, 28, 29, 31, 35, 36, 38, 43, 44, 45, 47, 48, 49, 54, 61, 63, 65, 68, 69, 72, 73, 74, 75, 78, 80, 83, 84, 86, 87, 88, 93, 98

Order book, 51, 52

Other type of museum (see "Museums"), 66

Outcomes, 87, 70

Output, 9, 10, 16, 45, 53, 68, 70, 80, 84, 87

Partners, 49, 71, 73, 74, 76

Pathfinder, 15, 17, 22, 33

Patrimoine numérique, 25

PDF, 57, 79

Photographs, 8, 17, 57, 65

KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

Plans, 3, 14, 46, 47, 48, 64, 65, 73, 81, 88

PNG, 57

Poland, 21

Portugal, 21

Postcards, 17, 65

Posters, 17, 65

Printed material, 17, 65

Progress, 1, 3, 6, 7, 8, 10, 19, 36, 40, 43, 51, 52, 53, 54, 57, 68, 73, 76, 77, 79, 81, 84, 85, 86, 96

Projection method, 66

Public access (see "Access"), 71

Public library (see "Library"), 66

Publishers, 10

Quality, 10, 33, 37, 39

Questionnaire, 3, 11, 14, 15, 16, 18, 19, 20, 21, 22, 24, 27, 33, 34, 35, 39, 40, 41, 42, 43, 45, 53, 57, 60, 61, 63, 71, 75, 76, 78, 79, 82, 83, 84, 85, 86, 87, 89, 97

Regional, 14, 22, 23, 24, 25, 26, 27, 28, 49, 52, 66, 73, 81

Relevant, 3, 6, 8, 14, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 38, 42, 44, 51, 57, 66, 67, 68, 69, 76, 77, 78, 79, 84, 85, 89, 93, 95, 96, 97

Research, 6, 8, 9, 11, 15, 18, 19, , 21, 22, 27, 38, 72, 74, 75, 76, 77, 86

Romania, 21, 83

Sample, 3, 13, 14, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 33, 35, 36, 38, 66, 75, 78, 79, 89, 93, 95, 97

Serials, 17, 65

SGML, 57

Slovak Republic (Slovakia), 21, 33, 83

Slovenia, 21, 83

SMIL, 58

Spain, 21, 29

Special interest group, 93, 96, 97

KEY WORD INDEX

(TABLE REFERENCES ARE OMITTED – PLEASE REFER TO THE LIST OF TABLES)

Specimens, 17, 65

Staff, 3, 15, 25, 39, 44, 45, 46, 68, 69, 74, 79, 80, 88

Statistical agencies (see "Central statistical offices"), 10

Statistical framework, 1, 7, 8, 9, 18, 75, 77, 78

Sweden, 21, 83

Text, 3, 38, 55, 57, 59, 84

TIFF, 57

Tourism, 9

Training for Audiovisual Preservation in Europe (TAPE), 7

TXT, 57

Unit cost, 15, 41, 58, 59

United Kingdom, 1

United States of America, 7

User access (see "Access"), 1, 11, 18, 42, 63, 82, 83, 87

Users, 17, 68

VHS, 57

Visitors, 17

Volumes, 16, 17, 54, 65, 83

Walloon, 24

WAVE, 58

Web (see "Internet"), 20, 79

Weighting, 3, 36, 37, 54, 66, 67, 68

Word (Microsoft file format), 2, 57

Works of art, 17, 65