

NUMERIC



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

# Pathfinder Survey Results

Summary analysis of a survey to test definitions, terms and ways of measuring digitisation activities in cultural institutions in Europe.

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

IPF Ltd  
No 1 Croydon  
CR0 0XT  
United Kingdom

**IPF**

5<sup>th</sup> December, 2007

<http://www.NUMERIC.ws>

## Emerging Pathfinder Survey Results

### Contents

Section		Page
1	Objectives	1
2	Survey Response	2
3	Results	2
4	Developing the Framework	9
Appendix		
I	Survey Questionnaire	13
II	List of Responders	18
III	Survey Definitions	19
IV	End Notes	20

#### Foreword:

The survey upon which this report is based is still on-going, although further responses will simply provide more reliable and representative results upon which to base modelling assumptions for estimating global activity in Europe. In the meantime, our thanks are extended to those international professional organisations and national ministries for their support in encouraging 60 individual institutions to participate in the survey.

The term "Pathfinder" has been applied to the survey because it was intended to test definitions and the practicability of questions, so as to refine a suitable questionnaire to employ in a series of sample surveys in 2008.

Given this background, **the following analysis should be understood to be purely illustrative of the position in a limited number of institutions.** The summary statistics are mostly un-weighted. i.e. the data for large and small institutions are treated equally, when calculating averages. Summaries are provided for each domain (Audio-visual / Film Institutes, Archives / Records Offices, Libraries, Museums), but the coverage was not sufficiently representative to provide measures for individual countries.

*In addition to the statistics, the survey also collated details relating to individual digitisation projects. These are not analysed as part of this report. A further model is employed in the study to estimate the overall digitisation activity in Europe, based on assumptions informed from the survey. Similarly, these estimates are not included in this report.*

## 1 OBJECTIVES

- 1.1 The European Commission<sup>1</sup> has adopted a recommendation to “gather information about current and planned digitisation of cultural material”. The Numeric study aims to establish a statistical framework to implement this recommendation. In considering the possible approaches, a pathfinder survey was conducted to evaluate definitions and practicable questionnaire designs with the intention of recommending a framework that balances the need for policy information with the effort of individual ministries and institutions to provide relevant data. The recommendations are contained in a report currently being considered by the European Commission.
- 1.2 The questionnaire was designed after benefiting from the experience of two previous surveys undertaken by the European Group of Museums (EGMUS) and the Council of National Librarians (CENL). The former is exploring the important aspect of relating digitisation activities to the size of the analogue collections held by their institutions. The latter has identified the need to relate practical measures of digitised output to the way in which the statistics about the analogue collections are held. For instance, “Volumes” is the analogue definition for books adopted under the ISO standards for library statistics, but the more convenient way of recording digitisation outputs is in “pages”. Therefore one objective of the questionnaire was to collect sufficient data to establish standard measures of analogue input to digitised output: How many pages per book, per linear metre of archive shelf space, images in a photograph collection, per hour of film, etc..
- 1.3 Another important objective of the pathfinder survey was to establish whether there are significant differences in the digitisation work undertaken by institutions serving national interests (that are relatively easier to survey) and those more numerous institutions that are dispersed to serve regional / local interests. Measures of difference inform assumptions about our estimates of pan-European activity.
- 1.4 The survey questionnaire is reproduced at appendix I, and was launched in September 2007 with the assistance of the Ligue des Bibliothèques Européennes de Recherche (LIBER), The Consortium of European Research Libraries (CERL) and the Association des Cinémathèques Européennes (ACE). In addition, professional and ministerial sponsorship was given to the exercise in Hungary, Portugal and Sweden. The data so collected is being made available to all sponsoring organisations.

---

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

## 2 SURVEY RESPONSE

2.1 Responses (as at 5<sup>th</sup> December, 2007) had been received from 60 institutions. The majority (45), occupy national positions, and the remainder (15) serve regional or local interests. The domains represented by each were reported as follows:

8	Archives	7	A-V / Film institutes
29	Libraries	13	Museums
2	Other types of institution <sup>2</sup>	1	Ministry <sup>3</sup>

2.2 The response at this point in time was also heavily weighted towards the active participation of Hungary, Portugal and Sweden in Europe, as the following shows:

2	Austria	1	Luxembourg (Grand-Duché)
1	Czech Republic	3	Netherlands
1	Denmark	1	Norway
1	Estonia	3	Poland
1	Finland	<b>14</b>	<b>Portugal</b>
2	France	<b>10</b>	<b>Sweden</b>
4	Germany	3	United Kingdom
<b>7</b>	<b>Hungary</b>	1	Turkey
3	Italy	1	Russian Federation

## 3 RESULTS

3.1 The primary purpose of the survey was to examine the practicability of collating data to certain definitions and to gather feedback in order to refine the approach to better coincide with the way in which institutions are able to provide consistent data. The lessons so learned are reported in the next section, "**FINDINGS FOR DEVELOPING THE FRAMEWORK**". However, so long as the analysis of the limited response is treated with extreme caution, there are a number of potentially interesting features and patterns worth noting, but it must be stressed that the following results are reported in this cautionary light.

### The scale of the digitisation agenda

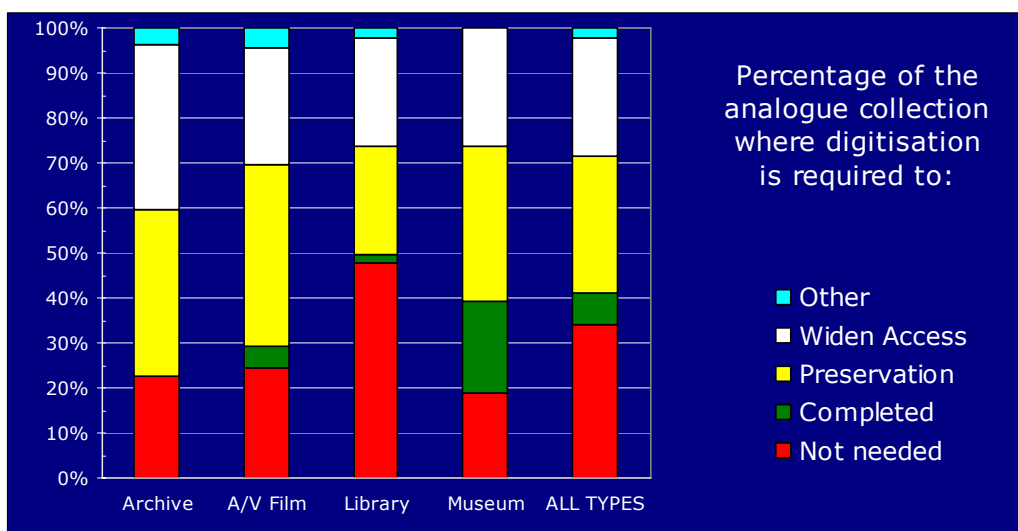
3.2 A significant number of respondents (22%) reported that all or most of their analogue collections needed to be digitised; this puts into proper context the overall survey average, where about one-third of the collections were reported as not needing to be digitised. Just under one half of the library collections did not need to be digitised (48% and the highest proportion amongst all the domains), but

<sup>2</sup> Reclassified as a Museum for the purpose of the analysis.

<sup>3</sup> Reclassified as a Library for the purpose of the analysis.

these libraries also reported that they had so far only succeeded in digitising less than 2% of their collections. The least progress appeared to be amongst Archives – (but this is weighted heavily by returns from 4 District Archives in one country: Portugal).

What percentage of your collection(s)	Archive / Records Office	Institute / Audio-visual or Film	Library	Museum	Grand Total
does not need to be digitised?	22.4%	24.3%	47.9%	19.0%	34.1%
has already been digitised?	0.2%	4.9%	1.9%	20.5%	7.1%
still needs to be digitised (preservation)?	36.9%	40.6%	23.8%	34.3%	30.4%
still needs to be digitised (widen access)?	36.7%	25.8%	24.4%	26.2%	26.4%
Still needs to be digitised for other reasons?	3.7%	4.4%	2.1%	0.0%	2.0%
Number of institutions	6	9	30	15	60

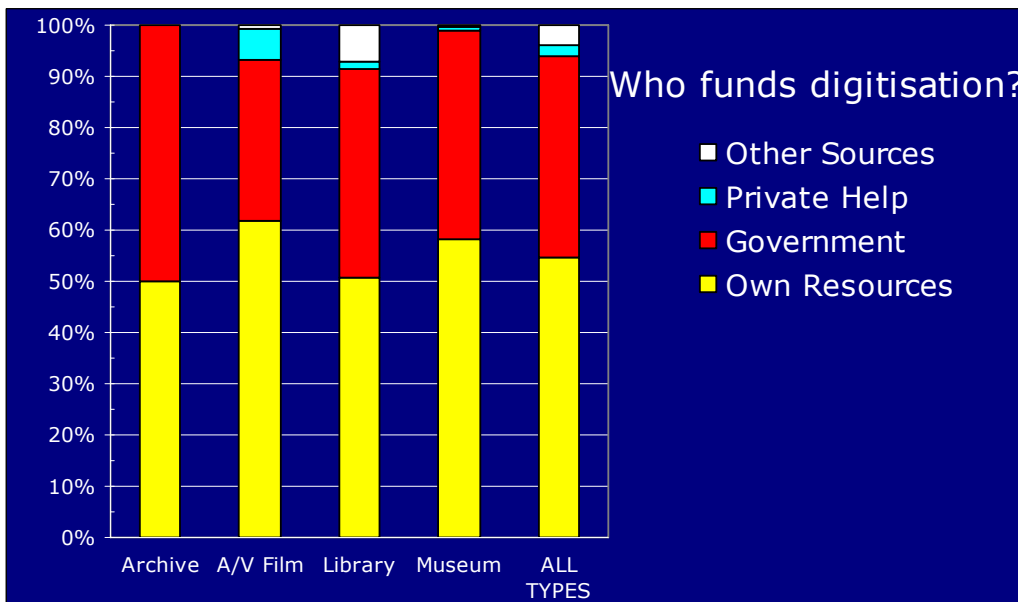
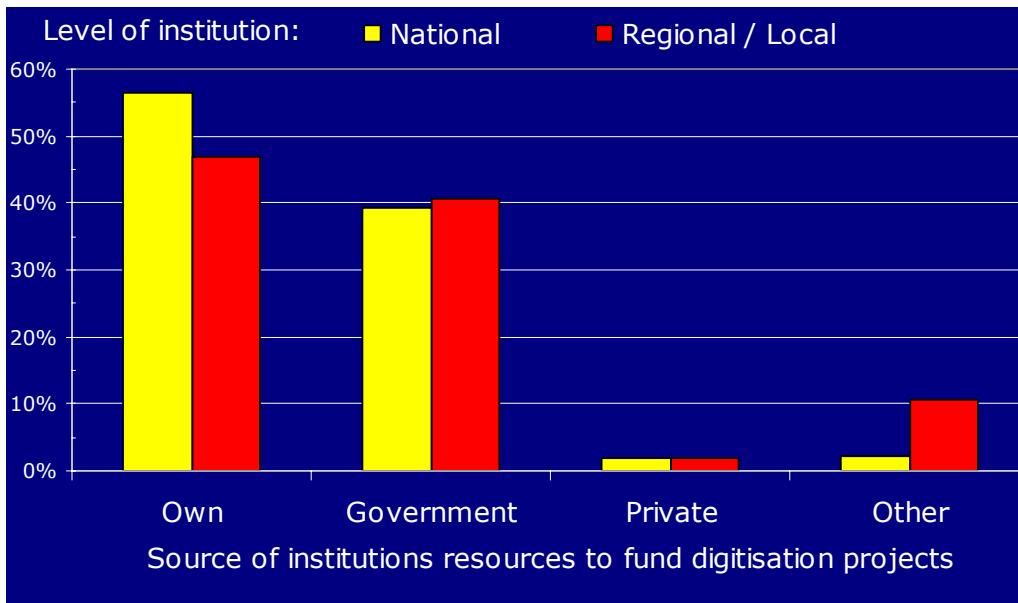


3.3 In practice many institutions were unable to differentiate between the need to digitise materials for preservation purposes, as distinct from the potential to widen their accessibility. This was epitomised by one respondent commenting that “if you could not find an item on the web, it did not exist.” Regardless, amongst all the domains, it is clear that the proportion of the collections that remain to be digitised (59%) is significantly more than the 7% that is claimed to have been already converted into electronic format.

### Finding the resources

3.4 More than half the funding of digitisation projects is provided by institutions from their own resources (55%), and central government assistance varies considerably, but is equivalent to just under 40%. The proportion of central government subvention made to institutions of regional / local importance, is similar to

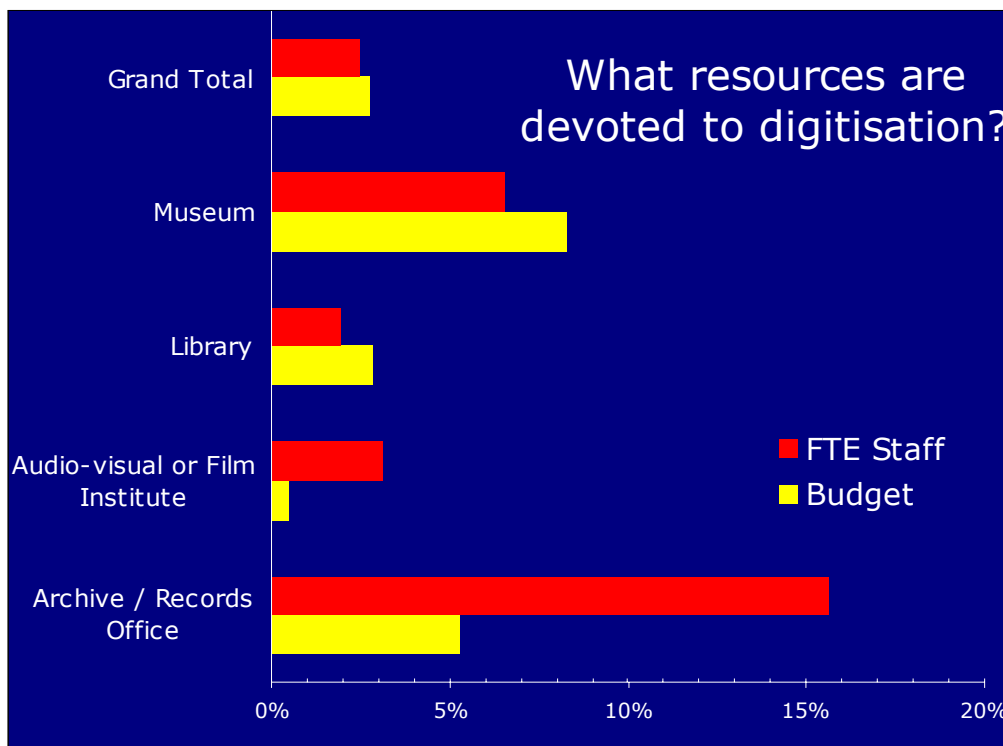
institutions of national importance / prominence suggesting that the digitisation agenda is recognised across all.



3.5 Regardless, it is clear that the significant proportion of project funding derives from internal funds, but although several institutions made the point that digitisation was not their most important priority, it is surprising that only about one-in-three have made formal digitisation plans or strategies:

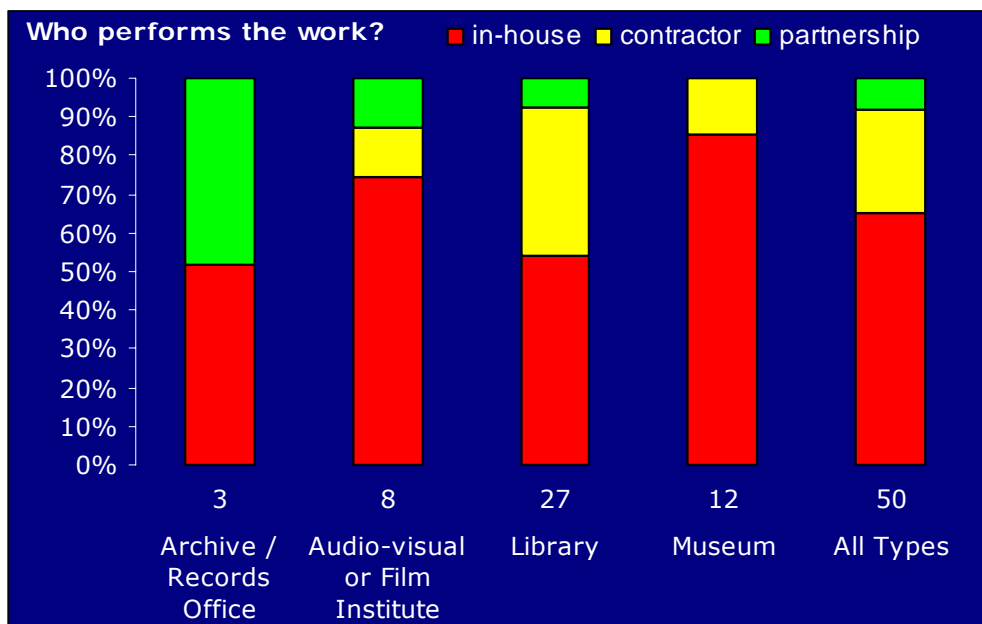
Do you have a digitisation plan?	Yes	No
Archive / Records Office	33%	67%
Audio-visual or Film Institute	33%	67%
Library	28%	72%
Museum	33%	67%
<b>All Types</b>	<b>31%</b>	<b>69%</b>

3.6 One reason, reported by several respondents, was the difficulty of either planning or setting identified digitisation budgets when the funding sources remained uncertain. Clearly, there are many competing demands on the revenue requirements of all institutions, and even where the digitisation agenda has been identified, the impression gained is one where many are having to make the best of marginal opportunities. Even so, the resources that are being devoted to digitisation projects are significant, even if a small proportion of the total budgets and staffing establishments. Excluding Archives, where a relatively high proportion of in-house staff in one institution, in particular, have been dedicated to digitisation projects, it is noticeable that the budget proportions are higher amongst museums.



**Acting alone, in consort or buying-in**

3.7 Most institutions (65%) depend upon in-house skills and resources to digitise their materials. Not all survey respondents (c.5/6<sup>ths</sup>) answered this question, but it is evident that large scale projects, that would benefit from the efficiencies that a contractor could provide, are more likely to arise in digitising text (printed or manuscript), and these sorts of projects are being pursued by libraries. The Film Institutes, in particular have specific and specialised needs, and for this reason may be better placed to manage the conversion processes themselves or working closely with their contractors.



3.8 A minority of institutions reported their intentions to either let external contracts or take on in-house temporary staff. There were other instances where a “leader” institution, provided both digitisation services and expertise to other “satellite” institutions, and in this context the more extensive partnership arrangements existed in the libraries domain.

**Standards and costs**

3.9 Many institutions found it very difficult to report the cost of digitisation work, for a number of reasons; see section 4. There was a considerable range, although some uniformity in the standards adopted.

3.10 *Text and Image:*

Resolution (dpi)	Min 72 : Median 300 : Max 4,800
Master File Formats	68% TIFF; 26% JPEG; 6% Other
OCR of digital output	21%
Cost per page for OCR	Min €0.25 Avg €1.13 Max €3.70

3.11 *Text Only:*

Resolution (dpi)	Min 200 : Median 300 : Max 1,200
Master File Formats	62% TIFF; 23% JPEG; 15% Other
OCR of digital output	35%
Cost per page for OCR	Min €0.40 Avg €0.93 Max €1.33

3.12 *Audio Material:*

Master File Formats	54% WAVE; 15% Broadcast WAVE; 31% Other
Cost per hour	Min €0.60    Avg €21.20    Max €71.84

3.13 *Video and Film:*

Master File Formats	23% AVI; 31% MPEG-4; 46% Other
Cost per hour	Min €5    Avg €398    Max €1,580

**Progress and plans**

- 3.14 The survey was designed to explore the overall progress made towards digitising the collections of the responding institutions. The main types of material likely to be held in collections were listed and the unit defined by which each analogue item should be counted. These mainly coincided with the ISO standards for statistical definitions in the cultural sector. The units adopted for each measure are defined on the second page of the questionnaire (see appendix 1).
- 3.15 Against the measures of total materials in the collection, respondents were asked to state how many represented "priority units". These were defined as materials requiring digitisation to fulfil preservation aims or widen access. Then, respondents were requested to provide the number of "achieved units": i.e. a figure for the materials that had already been digitised. A further request was made for "planned units", which were materials for which financial resources for digitisation had been identified in plans.
- 3.16 Information was also requested on the planned expenditure for digitising these materials, so that unit costs could be derived from the analysis. The equivalent number of pages was requested per "volume" of printed materials and "metre of archive shelving", in order to gain more consistent unit costs.
- 3.17 The overall distribution of responses is summarised in the following table. This shows the "Median", or mid-point in the distribution of responses. Some care is required in interpreting these summary figures as it is clear that in several instances the response values do not coincide with the questionnaire definitions. For instance, newspapers are often recorded as pages, issues, or titles, rather than volumes. Also the definition of prioritisation of materials for digitisation has been variously interpreted. ("*...*" denotes no valid data, or in some cases masking implausible results.)

		Weighted averages:				
		7.1%	43.5%	572	€ 1.24	
		Grossed Ratios - Median Values				
Objects / items / materials in collections:	Measure: [1]	Digitised as % of Collection	Priority as % of Collection	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
Books (including "Rare books")	Volumes	0%	11%	250	€ 156.64	€ 0.45
Newspapers	Volumes	2%	7%	14	...	€ 0.91
Journals and other serials	Volumes	0%	25%	350	€ 15.00	€ 0.30
Government publications	Volumes	16%	80%	723	€ 2,961.05	€ 3.72
Other type printed material	Number	6%	100%	3	€ 14.35	€ 3.70
Manuscripts	Number	2%	37%	45	€ 44.00	€ 8.74
Maps	Number	1%	13%		€ 10.75	
Photographs	Number	5%	30%		€ 4.07	
Engravings	Number	4%	71%		€ 19.94	
Drawings	Number	20%	100%		€ 4.82	
Posters	Number	17%	75%		€ 16.59	
Postcards	Number	44%	100%		€ 3.14	
Sheet music	Scores	1%	30%	23	€ 29.28	€ 0.68
Other images not classified above	Number	18%	80%		€ 5.13	
Archived records of government	Metres	50%	75%	768	...	€ 0.74
Archived records of historic importance	Metres	25%	47%	300	€ 398.86	€ 0.80
All other archived records	Metres	54%	100%	1,868	€ 2,900.73	€ 0.80
Man-made artefacts in museums	Artefacts	53%	86%		€ 3.89	
Natural world specimens	Objects	...	100%		...	
Works of Art - 2 dimensions	Exhibits	72%	81%		€ 10.51	
Works of Art - 3 dimensions	Exhibits	75%	100%		€ 10.51	
Other objects in museum collections	Objects	9%	100%		...	
Film and video recordings	Hours	4%	36%		€ 1,125.00	
Music and other recorded sound	Hours	9%	73%		€ 14.03	
Other items not classified above	Number	45%	...		...	

3.18 See pages 22:27 (appendix 4) for further information about the variation in these statistics.

3.19 Clearly the computed unit costs vary considerably and these compare with some indicative commercial rates<sup>4</sup> as follows:

- \* Books (bounded pages), depending on the volume, but much more on their state, the price per page varies €0.07 - €0.25.
- \* Newspapers depending on the volume and again much more on their condition, the price per page varies €0.08 - €0.35. However, for scanning + high structuring of newspapers, €1 is more realistic.

<sup>4</sup> These rates reflect estimated prices assuming high volume processing, and no assumptions concerning special requirements.

## Access to digitised materials

Main policy relating to usage and access to digitised materials / unique visitors / users per day:		Archive / Records Office	Audio-visual or Film Institute	Library	Museum	Grand Total
Free to full service	Nº Institutions	-	1	11	3	15
	Avg users/day	-	30	9,459	101	6,959
Paid to full or part	Nº Institutions	1	-	1	-	2
	Avg users/day	1	-	100	-	51
Restricted (e.g. in-house)	Nº Institutions	-	-	2	2	4
	Avg users/day	-	-	4,283	30	2,156
Something else <sup>1</sup>	Nº Institutions	-	1	1	1	3
	Avg users/day	-	130	-	1,100	615
<b>Total</b>	<b>Nº Institutions</b>	<b>1</b>	<b>2</b>	<b>14</b>	<b>6</b>	<b>23</b>
	<b>Avg users/day</b>	<b>1</b>	<b>80</b>	<b>8,051</b>	<b>244</b>	<b>4,971</b>

Note: <sup>1</sup> Free, paid or restricted depending on rights and use

3.20 As demonstrated in the table above, the prevalent policy adopted by institutions provided for free access to digitised materials. However, it is noticeable that the number of responders is comparatively low, raising some doubts as to whether or not formal policies are in always in place. Furthermore, the figures include a mix of both physical and remote access counts although, as might be expected, the average number of users is higher in those institutions with an open access policy.

## 4. DEVELOPING THE FRAMEWORK

### Financial statistics

4.1 The survey was designed to attempt to collect information about the costs in two ways: Directly seeking unit costs according to specific standards; requesting information about the volume of planned work and the total expected cost, from which unit costs could be computed when analysing the data. However, it is evident from the survey results that this proved the most problematic area for most respondents, and this was reflected in the paucity of information provided in the returned questionnaires. Comments provided also raised the need to define exactly what should be included in the cost, and in particular, whether or not it should include a recharge for staff time. For consistency's sake, *staff time should be recharged to the reported expenditure in future surveys.*

4.2 Furthermore, it was suggested that costing information was commercially sensitive. However, the consensus of views, shared by external contractors we have consulted, is that the cost varies considerably according to the standards sought, volume of work, and way in which the project is organised. The publication of illustrative benchmarks can only serve to better inform local

judgements concerning both the value and relative complexity of planned projects. *Therefore, this sort of qualitative detail will be required from the framework.*

- 4.3 Some respondents pointed to the difficulty of costing plans because the estimates often proved to be very wide of the actuality. Others stated that it was more difficult to trace the cost of work completed, because the accounting information was unavailable. However, the digitisation agenda is so substantial and the investment required so significant that it is evident that *more importance needs to be given to both planning and accounting for the activity, both within the institution and amongst sponsoring ministries.*

#### **Input – Output measures**

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

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

#### **Framework design**

- 4.5 The concept of prioritising materials for digitisation appears to be a philosophical one. A better definition is required to identify what proportion of the analogue collection coincides with a completed digitisation programme. The most prevalent approach taken by institutions is to work their way through the digitisation of specific collections. This defines their priorities, and the available budgets (whether notional or formally earmarked) mark the speed at which these projects progress.
- 4.6 *The framework needs to explore the stages in the digitisation process.* Partial digitisation can commence with the catalogues, prior to the conversion of the materials to which these refer. This can be catered for by including a measure covering the proportion of the catalogue available on-line (restricted or open access).

- 4.7 Another stage in the process relates to the proportion of materials in image format awaiting optical character recognition. A further category relates to the proportion of materials in primitive or now defunct digital formats, that are being converted to higher quality or modern standards.
- 4.8 It is apparent that making an artificial *distinction between digitising materials for preservation purposes and to widen their accessibility serves no statistical purpose.*
- 4.9 A number of further suggestions and comments have been made by survey respondents that will be considered. However, there is a simple balance to be struck – between simplicity (without which the survey response rate will be prejudiced), and more specific detail (which is required to add value to the analysis). The study approach employs deriving survey statistics to inform assumptions that are made in estimating the global activity in each country and domain.
- 4.10 To provide some structure to our consultations, a policy pro-forma was made available on the Numeric web-site. We invited ministries, professional associations, institutions and other interested stakeholders to provide views on the sort of information the statistical framework should provide. We considered it important to seek such views, rather than depend solely on reviewing prior studies, many of which had more restricted ambitions than developing a cross-domain / pan-European framework.
- 4.11 **The overwhelming message provided was the need to establish the link between the investment in digitisation projects and the benefit to those persons who would gain access to the materials.** “There’s a lack of information about the access and the use of digitised content by the public; having this sort of information could really address and support the future of digitisation initiatives”, was a typical response.
- 4.12 Another angle on the same point was expressed as follows: “There has been significant investment into digitisation. Unfortunately, I am still to be convinced that what has been digitised is being used to its capacity.” The importance of gaining hard facts to inform funding policy decisions was recognised by everyone.
- 4.13 Consultees said that many issues were equally important, and ranking them was therefore impossible, but preference was shown for information about the “Cost” and “Access / Usage” of digitised materials, followed by statistics about the progress “Achieved” and “Planned”. To an extent, information about the “Formats” was regarded as contextual rather than of strategic importance for the policy debate.

- 4.14 "Funding" was reported as the single most problematic issue relating to digitisation projects, "Copyright" issues were also significant. The point was also made that, once the materials had been digitised, further investment was needed to present the information to users. However, the positive impact that digitisation programmes were making in benefiting future generations and opening up access now to the cultural heritage for all was stressed.
- 4.15 The worth of developing the statistical framework is recognised – but unless institutions respond to future surveys, the statistics that are capable of influencing policy will never be generated. This really is an instance where actions count more than words! It is equally clear that unless both ministries and institutions believe that their efforts to generate the data will really stimulate funding flows, the actions will not happen.

A more detailed exemplification of the statistical analysis is shown at appendix 4



/  
15 Please provide your best estimate of the materials in your collections and digitisation activity

ONLY INPUT DATA IN CELLS CORRESPONDING TO TYPES OF MATERIAL HELD BY YOUR INSTITUTION (Use the yellow spaces)

[Click here for definitions](#)

Objects / items / materials in collections:

		Analogue collection			Digital output		
		Measure: [1]	total units: [2]	"priority" units: [3]	achieved units: [4]	planned pages: [6]	cost €: [7]
a	Books (including "Rare books")	Volumes					
b	Newspapers	Volumes					
c	Journals and other serials	Volumes					
d	Government publications	Volumes					
e	Other type printed material not classified above	Number					
f	Manuscripts	Number					
g	Maps	Number					
h	Photographs	Number					
i	Engravings	Number					
j	Drawings	Number					
k	Posters	Number					
l	Postcards	Number					
m	Sheet music	Scores					
n	Other images not classified above	Number					
o	Archived records of government / administration	Metres					
p	Archived records of historic importance	Metres					
q	All other archived records	Metres					
r	Man-made artefacts in museums	Artefacts					
s	Natural world specimens	Objects					
t	Works of Art – 2 dimensions	Exhibits					
u	Works of Art – 3 dimensions	Exhibits					
v	Other objects in museum collections	Objects					
w	Film and video recordings	Hours					
x	Music and other recorded sound	Hours					
y	Other items not classified above	Number					

Report both analogue and digitised materials in the units defined in column 1. Show equivalent pages for printed matter in column 6.

"Priority units" are materials requiring digitisation to fulfil preservation aims or widen access.

"Achieved units" are materials that have already been digitised.

"Planned units" are materials for which financial resources for digitisation have been identified in plans.

"Scores" represent "physical units" of printed music.

16 What currency is represented under column [7]

**Description of digitisation activity in the past 12 months**

(/)

- 17 What percentage of your digitisation work is carried out:
  - i in-house?
  - ii by an external contractor?
  - iii by a partner institution?

Leave this box blank

**Digitisation of:**

- 18 Resolution in dpi
- 19 Most frequent File Format of Master Files:
- 20 What percentage of digital output is OCRed?
- 21 What is the average cost per page for OCR achieved?  
*(i.e. inclusive of digitisation cost)*
- 22 In what currency are you reporting Q21?

**text & images:**

**text only:**

Leave this box blank

Leave this box blank

**Digitisation of audio material:**

- 23 File Format of Master Files
- 24 What is the average cost per hour achieved?
- 25 In what currency are you reporting Q24?

Leave this box blank

**Digitisation of video and film:**

- 26 File Format of Master Files
- 27 What is the average cost per hour achieved?
- 28 In what currency are you reporting Q27?

Leave this box blank

**Usage of, and access to digitised materials:**

- 29 Is access mainly to full service free, paid or restricted?
- 30 How many users / unique visitors access digitised materials per day?

Leave this box blank

/

The table below should identify the quality and quantity of the **digital** content (and related metadata) held by your institution. Please enter details in the yellow cells and start a new row for each collection.

Heading	Type of material, description of collection	Quantity & Definition	Format & Quality	IPR	Current Use	Existing Metadata	Language	How can it be accessed? OAI/PMH or Z39.50?
Example entry	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Example entry	<i>Text, image, movie, sound, music etc and also description of the collection [e.g. Provincial newspapers 1800-1850; theatre posters 1900-1999]</i>	<i>1,000 film clips, 2 million pages, 20,000 books etc.</i>	<i>Format - JPEG, MPEG, Quicktime, HTML, PDF etc., Quality - Resolution, sampling rate, colour/greyscale etc.</i>	<i>Access rights to use the content e.g. public domain or license</i>	<i>Describe current users of the content, including the number (e.g. registered users)</i>	<i>Describe fields, languages and structure of the metadata</i>		
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								



## List of Responders

## Appendix 2

### Archive / Records Office

Arquivo Distrital de Beja  
 Arquivo Distrital de Castelo Branco  
 Arquivo Distrital de Évora  
 Arquivo Distrital de Leiria  
 Cinémathèque française  
 Haparanda stadsbildarkiv, Haparanda stadsbibliotek  
 Hungarian Radio Archive  
 Landsarkivet i Härnösand

### Audio-visual or Film Institute

Deutsche Kinemathek - Museum für Film und Fernsehen  
 Deutsches Filminstitut - DIF  
 Finnish Film Archive  
 Norwegian Film Institute  
 Svenska Filminstitutet  
 The Swedish National Archive of Recorded Sound and Moving Images

### Government Department / Ministry

National Office of Cultural Heritage, Hungary

### Library

Academic Library of Tallinn University  
 Bayerische Staatsbibliothek  
 Biblioteca Nazionale Centrale di Roma  
 Biblioteka Uniwersytecka we Wrocławiu  
 Bibliothèque nationale de Luxembourg  
 Bibliothèque Nationale et Universitaire de Strasbourg  
 Det Kongelige Bibliotek (The Royal Library, Copenhagen)  
 Deutsche Nationalbibliothek  
 Instituto Superior Miguel Torga  
 Jász-Nagykun-Szolnok Megyei Verseghy Ferenc Könyvtár és Művelődési Intézet  
 Koninklijke Bibliotheek  
 Kungl. Biblioteket/National Library of Sweden  
 National Library of Poland  
 National Library of Scotland  
 National Library of Wales (Reported as "Other Type of Institution" - see notes below)  
 National Széchényi Library  
 Ottoman Bank Archives and Research Centre Library  
 Oxford University Library Services  
 Poznan University Library, A.Mickiewicz University  
 Serviços de Documentação - Universidade de Aveiro  
 Serviços de Documentação - Universidade de Aveiro  
 The National Library of Russia  
 The research Library of South Bohemia  
 Trinity College Dublin  
 University Library Maastricht  
 Uppsala University Library  
 Utrecht University Library  
 Vienna University Library

### Museum

Alberto Sampaio Museum  
 Länsmuseet Gävleborg  
 Museo Nazionale del Cinema-Fondazione Maria Adriana Prolo  
 Museu da Terra de Miranda  
 Museu de Grão Vasco  
 Museu Dr. Joaquim Manso  
 Museu Nacional de Arte Antiga  
 Museu Nacional de Machado de Castro  
 Museu Nacional do Traje e da Moda  
 Museum of Ethnography  
 Museum of Fine Arts, Budapest  
 Museum of Literature Petőfi  
 Östergötlands länsmuseum  
 Swedish national collections of music  
 Västmanlands läns museum

### Other type of organisation

Institut für Realienkunde des Mittelalters und der frühen Neuzeit  
 University Libraries system, Italy

*Note: The institutions reported themselves under the above domains. However, for the purposes of the analysis, some institutions have been reclassified to one of the four major types according to their substantive collections.*

## Survey Definitions

## Appendix 3

The notes available to survey respondents were accessible from the NUMERIC website:

[http://www.numeric.ws/content/default.asp?catref=4&content\\_ref=23](http://www.numeric.ws/content/default.asp?catref=4&content_ref=23)

These definitions were drawn from the following sources:

CIDOC, International Committee for Documentation of the International Council of Museums (ICOM), International Guidelines for Museum Object Information: The CIDOC Information Categories, available at:

<http://www.willpowerinfo.myby.co.uk/cidoc/guide/guide.htm>

The International Council of Museums <http://icom.museum/>

ISO 2789 (2006) Information and documentation – International library statistics

ISO 5127 (2001) Information and documentation - Vocabulary

Jones, M. and Beagrie, N. (2003), Preservation management of digital materials: a handbook, London: British Library, available at:

<http://www.dpconline.org/graphics/intro/>

Museums Association, UK (1998), available at:

[http://www.museumsassociation.org/home&\\_IXPOS\\_mahead1](http://www.museumsassociation.org/home&_IXPOS_mahead1)

Reitz, J. M., ODLIS - Online Dictionary for Library and Information Science, available at: <http://lu.com/odlis/>

The Society of American Archivists - A Glossary of Archival and Records Terminology (<http://www.archivists.org/glossary/>) also provides a lot of useful terms (including those related to technology).

Verheul, Ingeborg (2006), Networking for digital preservation. Current practice in 15 National Libraries, IFLA-Publications 119, available at:

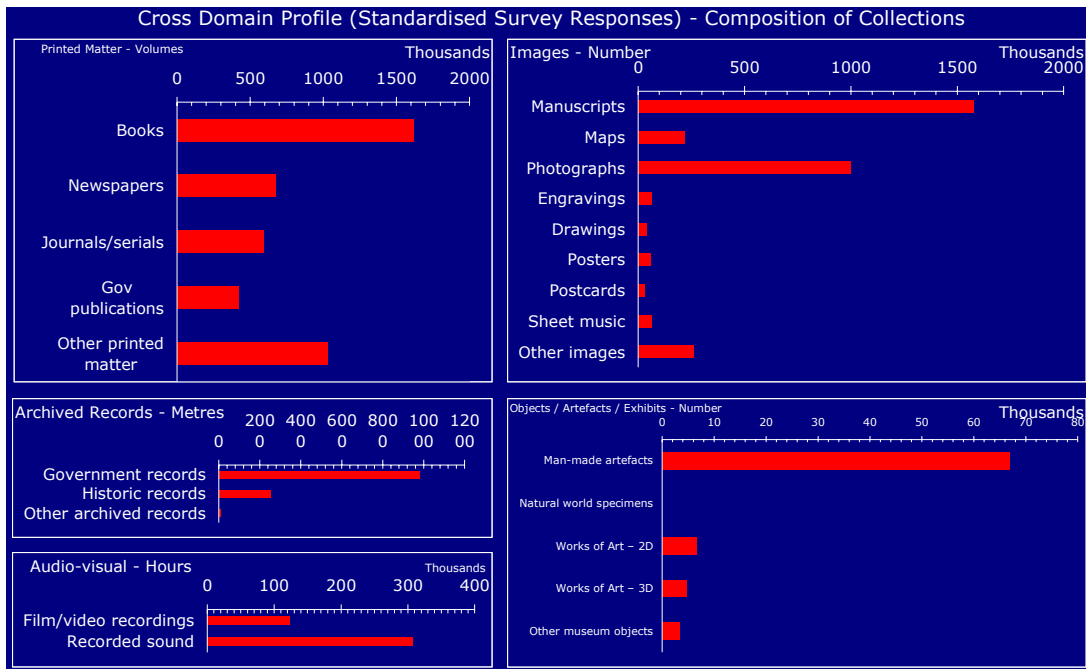
<http://www.ifla.org/VI/7/pub/IFLAPublication-No119.pdf>

**End notes**

**Appendix 4**

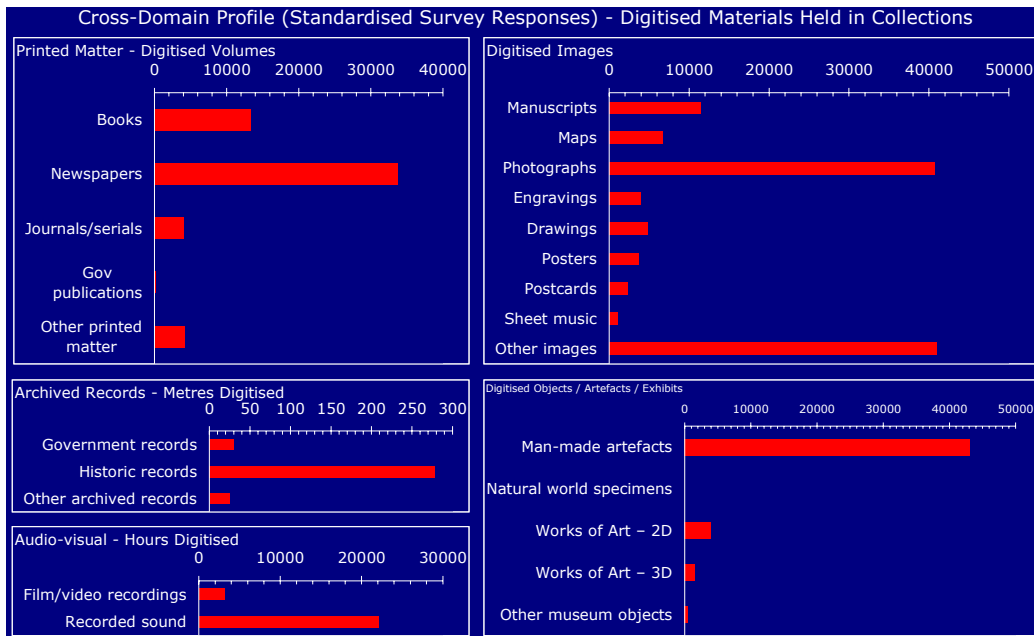
- IV.1 Rather more data than are available to this survey are required before a reliable commentary on overall EU27 digitisation activities can be reported. The response to the pathfinder survey simply provides a picture based on just 60 institutions, many of which are located in Hungary, Portugal and Sweden. Albeit biased towards this selection of institutions, a very broad-brush estimate of the annual EU spending devoted to digitisation would be about €370 millions, involving about 5,700 man years of in-house staff time.
- IV.2 In order to summarise the survey results covering different materials and measures across all the domains, the data in the following three summary charts have been standardised: That is an average for a notional all-domain institution has been calculated by dividing the totals by the number of valid responders. The analogue collections cover a very broad range of different materials, and it will be immediately apparent (from figure 1) that, because they are measured on different bases, it is difficult to compare the quantities of archived records of governments (measured in metres of shelf space occupied) with recorded sound (hours), or books (volumes), manuscripts and man-made artefacts.

Figure 1



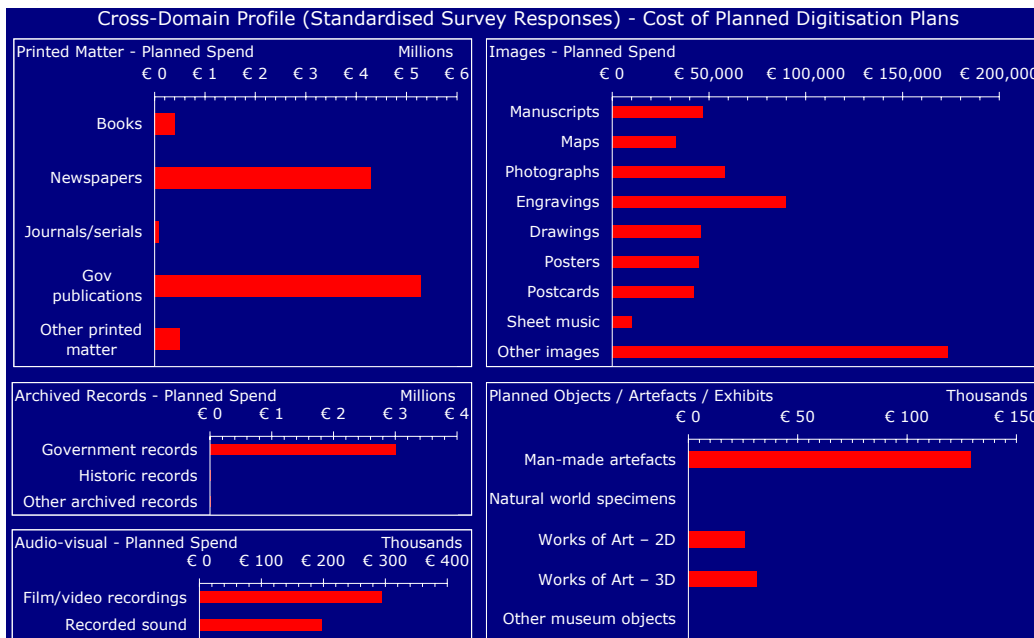
- IV.3 Figure 2 (overleaf) highlights how the digitisation activity has concentrated on newspapers, photographs and the conversion of historic records.

Figure 2



IV.4 The investment required to fund the digitisation plans is substantial, particularly for the printed matter held by libraries and archives; although, for the institutions in the other domains the smaller investment may represent larger slices of their available funds.

Figure 3



IV.5 The overall distribution of responses is summarised in the following tables. These show the minimum and maximum values provided by respondents, and the quarter-point values in between these ranges. The "Median", or mid-point in this distribution is denoted by the

value beside the "50%" mark ("..." denotes no valid data, or in some cases masking implausible results.)

Values at the points in the distribution	Books (inc Rare) - Volumes						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	0.0%	0.00%	0.0%	0.0%	1	€ 55
25%	1.5%	0.01%	0.3%	1.4%	172	€ 111	€ 0.28
<b>50%</b>	<b>10.9%</b>	<b>0.05%</b>	<b>8.5%</b>	<b>10.6%</b>	<b>250</b>	<b>€ 157</b>	<b>€ 0.45</b>
75%	41.7%	1.95%	16.1%	100.0%	369	€ 338	€ 1.31
Max	100.0%	97.1%	97.1%	100.0%	10,000	€ 5,000	€ 2.28
Valid Nº	24	24	16	12	11	8	8

Values at the points in the distribution	Newspapers - Volumes						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	0.0%	1.62%	24%	23.6%	14	...
25%	7.1%	1.67%	...	42.7%	14	...	€ 0.25
<b>50%</b>	<b>7.3%</b>	<b>1.73%</b>	...	<b>61.8%</b>	<b>14</b>	...	<b>€ 0.91</b>
75%	20.0%	50.86%	...	80.9%	14	...	€ 1.74
Max	100.0%	100.0%	...	100.0%	14	...	€ 2.28
Valid Nº	5	3	2	2	1	0	4

Values at the points in the distribution	Journals and other serials - Volumes						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	0.1%	0.00%	0.0%	4.0%	50	€ 1
25%	5.0%	0.02%	1.2%	16.0%	163	€ 8	€ 0.29
<b>50%</b>	<b>25.0%</b>	<b>0.10%</b>	<b>6.7%</b>	<b>60.0%</b>	<b>350</b>	<b>€ 15</b>	<b>€ 0.30</b>
75%	35.3%	2.65%	40.0%	...	1,875	€ 47	€ 0.40
Max	96.1%	100.0%	100.0%	...	6,000	€ 80	€ 2.28
Valid Nº	9	12	6	4	4	3	5

Values at the points in the distribution	Government publications - Volumes						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	25.0%	16.0%	18.2%	100.0%	400	€ 1,149
25%	52.5%	16.0%	18.2%	100.0%	561	€ 2,055	€ 3.30
<b>50%</b>	<b>80.0%</b>	<b>16.0%</b>	<b>18.2%</b>	<b>100.0%</b>	<b>723</b>	<b>€ 2,961</b>	<b>€ 3.72</b>
75%	84.0%	16.0%	18.2%	100.0%	884	€ 3,867	€ 4.14
Max	88.0%	16.0%	18.2%	100.0%	1,045	€ 4,773	€ 4.57
Valid Nº	3	1	1	1	2	2	2

Values at the points in the distribution	Other type printed material - Number						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	15.0%	2.99%	...	0.4%	2	€ 7
25%	20.0%	4.45%	...	0.4%	3	€ 10	€ 3.25
<b>50%</b>	<b>100.0%</b>	<b>5.91%</b>	...	<b>0.4%</b>	<b>3</b>	<b>€ 14</b>	<b>€ 3.70</b>
75%	100.0%	8.58%	...	0.4%	3	€ 18	€ 5.07
Max	100.0%	11.3%	...	0.4%	4	€ 22	€ 9.19
Valid Nº	5	3	0	1	2	2	4

Values at the points in the distribution	Manuscripts - Number						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	0.0%	0.00%	0.2%	0.1%	1	€ 1
25%	7.9%	0.17%	0.6%	0.8%	3	€ 9	€ 3.21
<b>50%</b>	<b>37.4%</b>	<b>1.81%</b>	<b>1.9%</b>	<b>11.1%</b>	<b>45</b>	<b>€ 44</b>	<b>€ 8.74</b>
75%	100.0%	15.70%	24.3%	100.0%	202	€ 323	€ 10.06
Max	100.0%	100.0%	...	100.0%	1,000	€ 1,812	€ 39.87
Valid Nº	18	19	14	9	7	5	5

Values at the points in the distribution	Maps - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	0.3%	0.0%	0.0%	3.1%
25%	3.4%	0.3%	2.6%	8.3%	€ 9.13
<b>50%</b>	<b>12.9%</b>	<b>1.1%</b>	<b>33.1%</b>	<b>25.0%</b>	<b>€ 10.75</b>
75%	62.5%	6.4%	45.0%	55.0%	€ 12.38
Max	100.0%	100.0%	60.0%	100.0%	€ 14.00
Valid Nº	12	12	7	4	2

Values at the points in the distribution	Photographs - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	0.1%	0.1%	1.2%	0.2%
25%	15.6%	1.1%	5.7%	17.5%	€ 3.00
<b>50%</b>	<b>30.0%</b>	<b>5.2%</b>	<b>15.7%</b>	<b>66.7%</b>	<b>€ 4.07</b>
75%	93.5%	28.2%	38.3%	100.0%	€ 4.66
Max	100.0%	100.0%	...	...	€ 6.00
Valid Nº	23	25	18	15	7

Values at the points in the distribution	Engravings - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	5.0%	0.2%	1.2%	25.0%
25%	9.2%	1.2%	10.0%	50.0%	€ 19.94
<b>50%</b>	<b>71.4%</b>	<b>4.4%</b>	<b>40.0%</b>	<b>90.0%</b>	<b>€ 19.94</b>
75%	100.0%	28.6%	52.5%	100.0%	€ 19.94
Max	100.0%	100.0%	95.0%	100.0%	€ 19.94
Valid Nº	7	9	5	5	1

Values at the points in the distribution	Drawings - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	9.5%	1.2%	1.2%	3.6%
25%	96.9%	8.6%	50.0%	5.0%	€ 3.84
<b>50%</b>	<b>100.0%</b>	<b>20.0%</b>	<b>87.5%</b>	<b>25.0%</b>	<b>€ 4.82</b>
75%	100.0%	85.7%	90.0%	100.0%	€ 9.82
Max	100.0%	90.0%	...	...	€ 22.43
Valid Nº	12	9	7	9	4

Values at the points in the distribution	Posters - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	10.0%	0.2%	1.3%	1.6%
25%	36.7%	4.6%	11.2%	1.8%	€ 8.32
<b>50%</b>	<b>75.0%</b>	<b>17.3%</b>	<b>48.9%</b>	<b>2.0%</b>	<b>€ 16.59</b>
75%	100.0%	31.0%	89.4%	51.0%	€ 24.86
Max	100.0%	100.0%	100.0%	100.0%	€ 33.13
Valid Nº	7	12	6	3	2

Values at the points in the distribution	Postcards - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	11.1%	0.1%	4.4%	50.0%
25%	50.0%	1.0%	76.1%	50.0%	€ 2.71
<b>50%</b>	<b>100.0%</b>	<b>44.3%</b>	<b>100.0%</b>	<b>50.0%</b>	<b>€ 3.14</b>
75%	100.0%	72.2%	...	50.0%	€ 3.58
Max	1000.0%	100.0%	...	50.0%	€ 4.01
Valid Nº	5	9	4	1	2

Values at the points in the distribution	Sheet music - Scores							
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page	
	Min	0.1%	0.0%	0.0%	0.2%	1	€ 1	€ 0.30
25%	4.5%	0.2%	2.4%	18.8%	12	€ 15	€ 0.30	
<b>50%</b>	<b>30.0%</b>	<b>1.3%</b>	<b>5.0%</b>	<b>62.5%</b>	<b>23</b>	<b>€ 29</b>	<b>€ 0.68</b>	
75%	100.0%	16.4%	78.3%	100.0%	47	€ 37	€ 1.18	
Max	100.0%	100.0%	...	100.0%	98	€ 45	€ 1.50	
Valid Nº	9	11	7	4	4	3	4	

Values at the points in the distribution	Other images - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	60.0%	0.2%	37.5%	100.0%
25%	70.0%	10.0%	53.1%	107.3%	€ 2.70
<b>50%</b>	<b>80.0%</b>	<b>17.9%</b>	<b>68.8%</b>	<b>114.6%</b>	<b>€ 5.13</b>
75%	90.0%	41.9%	84.4%	121.9%	€ 7.55
Max	100.0%	100.0%	100.0%	129.2%	€ 9.97
Valid Nº	2	4	2	2	2

Values at the points in the distribution	Archived govnt/adm records - Metres							
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page	
	Min	6.9%	0.1%	0.9%	100.0%	768	...	€ 0.74
25%	39.2%	25.0%	25.7%	100.0%	768	...	€ 0.74	
<b>50%</b>	<b>75.0%</b>	<b>50.0%</b>	<b>50.5%</b>	<b>100.0%</b>	<b>768</b>	<b>...</b>	<b>€ 0.74</b>	
75%	100.0%	75.0%	75.2%	100.0%	768	...	€ 0.74	
Max	100.0%	100.0%	100.0%	100.0%	768	...	€ 0.74	
Valid Nº	4	2	2	2	1	0	1	

Values at the points in the distribution	Archived records of hist import - Metres						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	0.9%	0.1%	2.5%	100.0%	100	€ 399
25%	22.0%	2.5%	4.8%	...	200	€ 399	€ 0.80
<b>50%</b>	<b>46.7%</b>	<b>24.9%</b>	<b>7.1%</b>	...	<b>300</b>	<b>€ 399</b>	<b>€ 0.80</b>
75%	75.0%	59.8%	53.6%	...	400	€ 399	€ 0.80
Max	100.0%	100.0%	100.0%	...	500	€ 399	€ 0.80
Valid Nº	9	4	3	2	2	1	1

Values at the points in the distribution	All other archived records - Metres						
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Pages per Unit	Cost per Planned Unit	Cost per Planned Page
	Min	32.0%	8.3%	8.3%	91.7%	100	€ 2,901
25%	66.0%	31.3%	31.3%	93.8%	984	€ 2,901	€ 0.80
<b>50%</b>	<b>100.0%</b>	<b>54.2%</b>	<b>54.2%</b>	<b>95.8%</b>	<b>1,868</b>	<b>€ 2,901</b>	<b>€ 0.80</b>
75%	100.0%	77.1%	77.1%	97.9%	2,752	€ 2,901	€ 0.80
Max	100.0%	100.0%	100.0%	100.0%	3,636	€ 2,901	€ 0.80
Valid Nº	3	2	2	2	2	1	1

Values at the points in the distribution	Man-made artefacts in museums				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	8.0%	1.4%	2.0%	7.1%
25%	24.7%	21.4%	28.1%	20.0%	€ 2.49
<b>50%</b>	<b>85.7%</b>	<b>53.1%</b>	<b>93.9%</b>	<b>30.0%</b>	<b>€ 3.89</b>
75%	100.0%	92.4%	...	100.0%	€ 5.28
Max	100.0%	95.0%	...	100.0%	€ 6.68
Valid Nº	6	7	6	5	2

Values at the points in the distribution	Natural world specimens				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	100.0%	...	...	...
25%	100.0%	...	...	...	...
<b>50%</b>	<b>100.0%</b>	...	...	...	...
75%	100.0%	...	...	...	...
Max	100.0%	...	...	...	...
Valid Nº	1	0	0	0	0

Values at the points in the distribution	Works of Art – 2 dimensions - Exhibits				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	20.0%	16.7%	20.6%	7.7%
25%	61.7%	36.2%	36.4%	9.1%	€ 5.80
<b>50%</b>	<b>81.0%</b>	<b>72.4%</b>	<b>92.3%</b>	<b>44.1%</b>	<b>€ 10.51</b>
75%	100.0%	89.2%	92.5%	100.0%	€ 15.22
Max	100.0%	92.5%	...	100.0%	€ 19.94
Valid Nº	7	6	5	5	2

Values at the points in the distribution	Works of Art – 3 dimensions - Exhibits				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	20.0%	13.3%	35.4%	25.0%
25%	53.2%	59.1%	65.1%	65.1%	€ 5.80
<b>50%</b>	<b>100.0%</b>	<b>75.0%</b>	<b>82.5%</b>	<b>89.2%</b>	<b>€ 10.51</b>
75%	100.0%	80.0%	...	100.0%	€ 15.22
Max	100.0%	90.0%	...	100.0%	€ 19.94
Valid Nº	6	5	4	4	2

Values at the points in the distribution	Other objects in museum collections				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	90.0%	8.3%	11.1%	50.0%
25%	95.0%	8.3%	20.8%	62.5%	...
<b>50%</b>	<b>100.0%</b>	<b>9.2%</b>	<b>30.6%</b>	<b>75.0%</b>	...
75%	100.0%	20.0%	40.3%	87.5%	...
Max	100.0%	50.0%	50.0%	100.0%	...
Valid Nº	3	4	2	2	0

Values at the points in the distribution	Film and video recordings - Hours				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	5.0%	0.0%	0.0%	0.0%
25%	10.9%	1.6%	5.0%	16.9%	€ 564
<b>50%</b>	<b>36.5%</b>	<b>3.8%</b>	<b>29.6%</b>	<b>40.0%</b>	<b>€ 1,125</b>
75%	72.5%	24.4%	47.3%	80.0%	€ 1,903
Max	100.0%	100.0%	1000.0%	100.0%	€ 3,113
Valid Nº	12	13	9	5	4

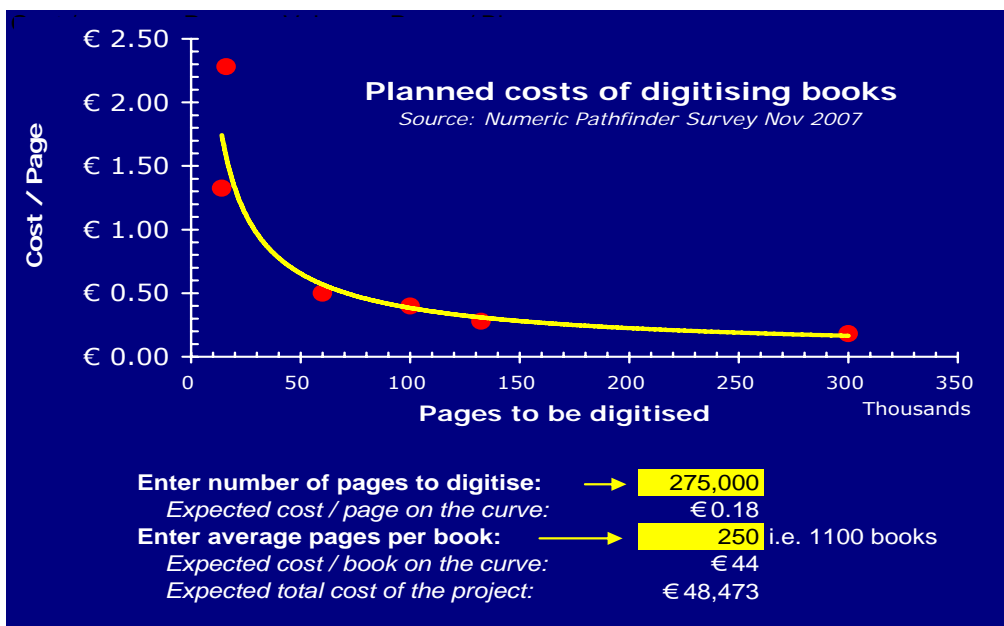
Values at the points in the distribution	Music and other recorded sound - Hours				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	1.0%	0.3%	0.3%	21.9%
25%	40.0%	5.5%	7.9%	25.0%	€ 7.32
<b>50%</b>	<b>72.7%</b>	<b>9.1%</b>	<b>12.4%</b>	<b>28.2%</b>	<b>€ 14.03</b>
75%	90.0%	30.9%	30.9%	31.3%	€ 82.02
Max	100.0%	100.0%	1000.0%	34.4%	€ 150.00
Valid Nº	7	5	5	2	3

Values at the points in the distribution	Other items - Number				
	Priority as % of Collection	Digitised as % of Collection	Digitised as % of Priority	Planned as % of Priority	Cost per Planned Unit
	Min	...	17.9%	...	100.0%
25%	...	31.3%	...	100.0%	...
<b>50%</b>	...	<b>44.6%</b>	...	<b>100.0%</b>	...
75%	...	58.0%	...	100.0%	...
Max	...	71.3%	...	100.0%	...
Valid Nº	0	2	0	1	0

IV.6 Brewster Kahle quotes US\$30 (about €44) as the cost of digitising a book included within the Internet Archive<sup>5</sup>.

IV.7 In Europe, the concept of a networked digital library is taking form in the European Digital Library. Obviously, the cost of digitising books varies according to many factors, such as the size of the books, quality of the image to be interpreted (for OCR), and the volume of work involved in the project. Based on the responses to the pathfinder survey, the comparable cost of digitising each book could vary, as the following estimator model shows, and to match the cost shown at IV.5 the project would need to digitise at least 275,000 pages:

IV.8 Figure 4



The cost can also vary according to whether or not any partners are able to cross-subsidise the work.

IV.9 Whilst overall support is well organised in the libraries domain, assistance is also provided by PrestoSpace to institutions with Audio-Visual interests. They have developed a questionnaire for gathering information about preservation practices and digitisation plans for audiovisual collections in Europe. The aim of the questionnaire is to ensure that the provided digital services and tools in the preservation factories fit the needs of the archives, service providers and end users. This provides a "Ready-reckoner" for estimating the resources required to process a range of different media from a broad range of carriers.

IV.10 The PrestoSpace and, indeed, the British Library LIFE costing model approach has several merits – the models provide excellent

<sup>5</sup> Open Content Alliance's annual meeting, October 17, 2007 in San Francisco

definitions, and the benchmark results provide a further incentive to complete the survey.

*Use of contractors and partnership arrangements*

IV.11 It was previously noted that specialised contractors were performing the digitisation work in relatively few institutions. Most outsourced work comprises larger scale projects which benefits from an industrial approach. Some institutions appear to scan the materials, and then pass these to a contractor to complete the OCR process. Even so, the list of contractors named by the limited number of institutions responding to the pathfinder survey suggests that there are many differing specialties being sought:

- \* Arcanum Adatbazis Inc.
- \* BBI
- \* Belser wissenschaftliches Dienst
- \* Carrena AB
- \* CCS GmbH
- \* Devo IT
- \* Eneclan
- \* GroKom Képm
- \* InduPrint Service GmbH,
- \* Infeurope SA,
- \* Infotechnique SA,
- \* Karmac; X-Cago
- \* mfm
- \* Microfilm Est = Herrmann & Kraemer (France)
- \* Microformat
- \* MIK-CENTER GmbH
- \* M&M Wandel GmbH
- \* Pictura
- \* PixelPrint,
- \* QBD - Quality by Design
- \* Softag Informática e Escritórios
- \* Stefan Steib
- \* Strata Preservation

IV.12 The support provided by many contractors offers assistance in kind in many instances, and the relationship is often reported as a partnership arrangement. Microsoft, Google, ProQuest, Octavo,

Digidata were quoted as such. Furthermore, a number of institutions have established the experience and technical expertise to act as digitisation hubs, and several universities and state archives fall into this category.