

The Cost of Digitising Europe's Cultural Heritage

A Report for the Comité des Sages of the European Commission

Prepared by Nick Poole, the Collections Trust

November 2010

References and Acknowledgements

We would like to acknowledge the support, guidance and expertise of the following people and organisations in the creation of this Study:

Marco de Niet, Digitaal Erfgoed Nederland
Gerhard jan Nauta, Digitaal Erfgoed Nederland
Gordon McKenna, Collections Trust
Yvo Volman, European Commission
Zofia Frahjer, European commission

We would like to thank the museums, libraries, archives and digitisation services that provided data in support of this research.

Contents

Executive Summary	1
Overview of Figures by Sector	5
1. Purpose of this Study	8
2. Defining Digitisation	11
3. Cost-Models for Digitisation	14
4. Understanding Formats	17
5. A Structured Approach to Normalising Costs	20
6. The Scale of European Cultural Heritage	23
7. The Cost of Digitising Libraries	25
8. The Cost of Digitising Museums	49
9. The Cost of Digitising Archives	58
10. The Cost of Digitising AV Material	65
11. The Cost of Digitising European Cultural Heritage	73
12. Comparisons	75
13. Methodology	77
14. Bibliography	78

Executive Summary

Introduction

Every year, Europe's museums, archives and libraries and Member State Governments expend considerable time, effort and resources to digitise their tremendously rich and diverse Collections, both to aid long-term preservation and to drive new models of public engagement.

The purpose of this study has been to conduct a thorough investigation of the costs associated with digitising different types of material in different types of cultural heritage institution. The objective has been to arrive at a set of reasonable projected costs for the Digitisation of Europe's cultural heritage.

It should be recognised from the outset that there can be no single definitive answer to this question. Digitisation is a process, and as with any process the actual cost depends both on the organisational context, the complexity of the material and the sophistication of the output.

That said, we have adopted as rigorous a methodology as possible in order to provide the Comité des Sages with the most robust projections possible to inform their recommendations. The following Executive Summary provides a brief overview of the key findings.

Overall Finding

This Report features completed data for the Digitisation of collections in Libraries, Museums, Archives and Audiovisual Archives. It does not include data concerning the broader AV collections held by Broadcasters, although we would recommend the inclusion of these in a future investigation.

- The estimated total cost of digitising the collections of Europe's museums, archives and libraries, including the audiovisual material they hold is approximately €100bn, or €10bn per annum for the next 10 years, factoring in a cumulative efficiency gain of 0.5% per annum.
- The cost of preserving and providing access to this material over a 10-year period after Digitisation would be in the order of €10bn to €25bn, provided that centralised repository infrastructure is made available for the purpose.
- The Research & Development Budget for the Joint Strike Fighter programme is estimated at €40.34bn.
- It would cost between 10% and 40% of the Joint Strike Fighter R&D budget to digitise every eligible title in Europe's libraries.
- The cost of delivering one Joint Strike Fighter is €147.41m, equivalent to the cost of digitising 1.93m books, or 2-3% of all individual titles held in libraries.
- The cost of delivering 100km of main road in Europe is €750m.
- 100km of main road is equivalent to the cost of digitising every piece of audio content in EU cultural institutions, or 48% of the total holdings of video (excluding film).
- 100km of main road would pay for the Digitisation of up to 16% of all available books in EU libraries, or 40% of the historic photographs in EU cultural institutions.

Findings Related to Libraries

- We estimate that, excluding multiples and series (for the purpose of clarity 'series' here refers to series of books, and does not include journals and newspapers, for which calculations are provided elsewhere), European Libraries hold a total of between 59 and 95 million individual book titles (mean 77 million).
- For comparison, the Google Book Search project has estimated the total number of individual titles *in existence* at 130m.
- The total number of pages to be digitised is approximately 1.47 to 2.36 billion (mean 1.92bn).
- The total *estimated mean* cost of digitising these books is between €4.79bn and €11.76bn (depending on the model adopted for the Digitisation).
- An investment of €100m will pay for the Digitisation of between 586,510 and 1,376,000 books (mean 981,255).
- The total book stock (including multiples and series) in European libraries is approximately 5.4bn books.
- Allowing for material that is too fragile, there are approximately 6.9 million rare books (including pamphlets and incunabula) to be digitised, at an estimated cost of €6.73bn to €10.51bn, dependent on the richness of the associated metadata (mean €8.62bn).
- The total cost of digitising all non-audiovisual material in European library collections is between €13.45bn to €30.89bn if the Digitisation is in-house or outsourced (mean €22.17bn) and €12.38bn to €21.01bn if it is done under a Public Private Partnership (mean €16.70bn).

Findings Related to Museums

- The 'eligible' collections of European museums (that is, collections that are not deemed too fragile to digitise) include 265m man-made artefacts and more than 221m natural objects.
- Digitising the eligible collections of European museums would cost between €13.75bn and €63.27bn (mean €38.51bn).
- The breadth of the range of potential costs indicates the breadth of material types covered by museum collections.
- The majority of museum Digitisation is completed as part of ongoing Collections Management & conservation activity.
- European museums house almost 75.43m individual works of art, including paintings, prints, drawings and sculpture.
- European museums house 350m photographs that are suitable for Digitisation, almost 20,000 photographs for every museum

Findings Related to Archives

- The National Libraries in the EU contain more than 26.98 billion pages of archival records, of which approximately 17.27 billion are eligible/appropriate for Digitisation.

- Digitising the eligible collections of the National Archives (including their branches and service points) in Europe would cost €41.87bn (mean average cost).
- There are approximately 692908 units of microfilm (accounting for countless more individual microfilm frames) in National Archives in the EU.

Findings Related to Audiovisual Collections

- There is no simple methodology for establishing costs-per-hour for the Digitisation of audiovisual collections
- On average, the cost of digitising film is 10x that of the cost of digitising video, because of the relative fragility of the material and the occurrence of non-standard formats and speeds.
- There are approximately 10.81 million hours of Audio material in European cultural institutions.
- There are approximately 12.14 million hours of Video in European cultural institutions.
- There are approximately 1.03 million hours of Film in European cultural institutions.
- The total cost of digitising the eligible AV material in European cultural institutions would be approximately €4.94bn.

Observations Arising from the Study

Several interesting points have arisen during the conduct of this study, which we present in this Summary for general information.

- Mass-Digitisation is an industrial process, and hence is very susceptible to efficiency gains at scale. Broadly, the larger a Digitisation project becomes, the lower the unit cost of Digitisation due to the dispersal of overhead and upfront capital costs over a larger body of material.
- Digitisation naturally tends towards greater efficiency over time. Hence the work-rate of a Digitisation facility at the start of a 10-year project is significantly lower than towards the end. This is not only due to the increased technical competence of the people involved, but also the gradual refinement and streamlining of workflow over the period of the project. This effect has been noted even in 3-year projects, with a significant majority of the Digitisation completed during the project lifetime being completed in the 3rd year.
- Due to natural market forces and the increasing presence of low-cost, high-volume Digitisation services in India and the Far East, the unit cost of Digitisation provided by 3rd party services has decreased significantly over the past 10 years.
- The Return on Investment in large-scale Digitisation tends to be higher where there is ongoing strategic investment in Digitisation as a core activity of the cultural institution, rather than as a project-funded activity. This is due to the significant efficiency gain that arises from long-term practice and the development of tacit institutional knowledge and expertise leading to more effective prioritisation and decision-making.
- As with physical collections, the acquisition of digital material creates a long-term obligation on the host institution, which must be accounted for. Most estimates put the cost of preserving and providing access to a digital asset for a period of 10 years at 50-100% of the initial costs of creating it. Hence, mass-Digitisation creates a large-scale economic obligation which must be addressed from the outset in programme budgets.

- When considering mass-Digitisation and text-conversion/encoding of books, it is important to factor into the value equation not only the potential value of the assets themselves, but also the considerable value of the new knowledge and information associated with them (such as the new insight into language development gained by Google as a result of the Google Book Search project).

Overview of Figures by Sector

Please note that the justification and/or margin of error for the figures reproduced below is contained elsewhere in the report. These figures should not be quoted out of the context in which they were calculated.

Libraries

<i>Description</i>	<i>Value</i>	<i>Units</i>
National Libraries in the Council of Europe	45	Sites
National Library service points	178	Service Points
Public Libraries	205336	Sites
HE/University Libraries	10161	Sites
School Libraries	164436	Sites
Special Libraries	29089	Sites
Google books estimated number of individual titles	130	Million titles
Lower estimate of titles currently in EU libraries	59	Million titles
Upper estimate of titles currently in EU libraries	95	Million titles
Mean estimate of titles currently in EU libraries	77	Million titles
Lower estimate of pages	1.47	Billion pages
Upper estimate of pages	2.36	Billion pages
Mean estimate of pages	1.93	Billion pages
Lower cost estimate(in-house Digitisation)	124	€ per book
Upper cost estimate (in-house Digitisation)	170.50	€ per book
Lower cost estimate (outsourced Digitisation)	73.15	€ per book
Upper cost estimate (outsourced Digitisation)	103.50	€ per book
Lower cost estimate (Public Private Partnership)	55.15	€ per book
Upper cost estimate (Public Private Partnership)	66.50	€ per book
Mean total cost for BOOK digitisation (in-house)	11.76	€bn
Mean total cost for BOOK digitisation (outsourced)	7.08	€bn
Mean total cost for BOOK digitisation (PPP)	4.79	€bn
Estimated number of RARE BOOKS	6.9	Million books
Estimated (mean) cost of RARE BOOK digitisation	8.62	€bn
Estimated (mean) cost of ARCHIVAL digitisation in libraries	2.45	€bn
Estimated pages of newspaper (based on 20m per MS)	540	Million pages
Estimated (mean) cost of NEWSPAPER digitisation in libraries	0.68	€bn
Estimated number of MAPS in library collections	7.23	Million maps
Estimated (mean) cost of MAP digitisation in libraries	61.03	€m
Estimated number of PHOTOGRAPHS in library collections	8.64	Million photos
Estimated (mean) cost of PHOTO digitisation in libraries	17.62	€m

TOTALS:

Estimated total cost of digitisation in LIBRARIES (in-house)	23.67	€bn
Estimated total cost of digitisation in LIBRARIES (outsourced)	18.95	€bn
Estimated total cost of digitisation in LIBRARIES (PPP)	16.70	€bn

Museums

<i>Description</i>	<i>Value</i>	<i>Units</i>
Total number of museums in the EU	17673	Sites
Total number of museums with an 'Art/History/Archaeology' focus	7776	Sites
Total number of museums with a 'Science/Technology' focus	6716	Sites
Total number of other types of museum	3181	Sites
Estimated number of MAN-MADE objects in museums	265	Million items
Estimated mean cost of digitising MAN-MADE objects in museums	15.34	€bn
Estimated number of NATURAL SPECIMEN in museums	307	Million items
Estimated number of NATURAL SPECIMEN suitable for digitisation	221	Million items
Estimated mean cost of digitising NATURAL objects in museums	16.23	€bn
Estimated number of WORKS OF ART in museums	75.43	Million items
Estimated mean cost of digitising WORKS OF ART in museums	4.38	€bn
Estimated number of PHOTOGRAPHS in museums	350	Million items
Estimated mean cost of digitising PHOTOGRAPHS in museums	2.78	€bn

TOTALS:

Estimated total cost of digitising MUSEUM COLLECTIONS in the EU	38.73	€bn
---	-------	-----

Archives

<i>Description</i>	<i>Value</i>	<i>Units</i>
Number of National Archives included in survey	25	
Total holdings of archival material	26.98	Billion pages
Total remaining to be digitised	10.45	Billion pages
Estimated total cost of digitisation of ARCHIVAL RECORDS	41.79	€bn
Estimated total holdings of MICROFORMS	692908	Units
Estimated total cost of digitising MICROFORMS	11.73	€m
Estimated total holdings of PHOTOGRAPHS	8.29	Million units
Estimated total cost of digitising PHOTOGRAPHS	64.51	€m

TOTALS:

Estimated total cost of digitising NATIONAL ARCHIVES in the EU	41.87	€bn
--	-------	-----

Audiovisual Archives

<i>Description</i>	<i>Value</i>	<i>Units</i>
Estimated total number of hours of AUDIO in EU cultural institutions	10.81	Million hours
Estimated total number of hours of VIDEO in EU cultural institutions	12.14	Million hours
Estimated total number of hours of FILM in EU cultural institutions	1.04	Million hours
Estimated total number of hours of AV in EU cultural institutions	23.99	Million hours
Estimated mean cost of digitising AUDIO	0.67	€bn
Estimated mean cost of digitising VIDEO	5.36	€bn

Estimated mean cost of digitising FILM	1.03	€bn
--	------	-----

TOTALS:

Estimated total cost of digitising AV MATERIAL in the EU	4.94	€bn
--	------	-----

OVERALL TOTALS

<i>Description</i>	<i>Value</i>	<i>Units</i>
Estimated MEAN cost of digitising LIBRARY COLLECTIONS in the EU	19.77	€bn
Estimated total cost of digitising MUSEUM COLLECTIONS in the EU	38.73	€bn
Estimated total cost of digitising NATIONAL ARCHIVES in the EU	41.87	€bn
Estimated total cost of digitising AV COLLECTIONS in the EU	4.94	€bn

Estimated total cost of digitising CULTURAL MATERIAL in the EU*	105.31	€bn
---	--------	-----

* NOTE This figure is subject to significant caveats in relation to scope, material included and the basis of calculation. Please refer to the wider report for context to this calculation.

1. Purpose of this Study

1.1 The European Context

- 1.1.1 Over the past 10 years, the European Commission and Member States have invested millions of Euro in supporting cultural heritage institutions (museums, archives and libraries) to digitise their Collections and make them available on the Internet (between 2005-2009 the European Commission invested €149m in the eContentPlus Programme alone).
- 1.1.2 This investment has driven an unprecedented period of technical Research & Development across the European cultural heritage sector, in which organisations of all scales have developed different models and methodologies for Digitisation.
- 1.1.3 Alongside the tremendous *public* investment in Digitisation, the past decade has seen the emergence of medium to large-scale programmes of Digitisation led by commercial enterprises such as Google, Proquest and Microsoft.
- 1.1.4 In turn, these investments have driven a secondary commercial and part-subsidised market of bespoke Digitisation and Digital Preservation services and consultancies.
- 1.1.5 The Digitisation of European cultural heritage has significantly improved the accessibility of this material for research, learning and enjoyment. Yet, at the same time, it has created concerns about sustainability, Return on Investment and long-term cost.

1.2 The Role of Europeana

- 1.2.1 Europeana (<http://www.europeana.eu>) is both a celebration and a consolidation of European cultural identity on the Web. By providing a single, multi-lingual point of access to digital cultural material, it acts as a gateway to Europe for a global audience.
- 1.2.2 In addition to its public role, Europeana has played an important function in bringing together the professional and technical cultural heritage communities and in driving innovation, collaboration and the creation of new content.
- 1.2.3 By providing a public channel for European digital cultural content, Europeana is driving improvements in the quality and integrity both of the content and of the knowledge which cultural heritage institutions make available alongside it.
- 1.2.4 In future, Europeana has the potential to make a significant contribution in opening up and democratising access to and engagement with Europe's cultural heritage – both directly through its public portal and indirectly in its role as a broker to the Media, Broadcasting, Publishing and other Creative Industries.
- 1.2.5 In order for Europeana to fulfil this potential, however, it depends on the sector achieving a scalable and sustainable model for the creation, management and distribution of new digital content.

1.3 The Comité des Sages

- 1.3.1 The Comité des Sages has been challenged to make recommendations to the Commission and to cultural actors, Governments and agencies throughout the EU concerning how best to capture, foster, share and celebrate the diversity and excitement of European culture and creativity online.

1.3.2 In particular, the Comité is addressing 3 areas:

- Funding sources for Digitisation
- Interaction between public and private organisations in the Digital age
- Solutions for Digitisation of public domain and in-copyright material

1.3.3 The question of establishing reliable, consistent cost-models for Digitisation spans all 3 of these lines of enquiry. It impacts on the scale and nature of the funding required, the potential for sustainable revenue models, and the value proposition for addressing collective and transactional rights clearance.

1.3.4 The purpose of this report is therefore primarily to provide the Comité des Sages with a reasonable framework within which both to advocate for solutions to these issues, and to make evidence-based recommendations to the European Commission for future investment in Digitisation.

1.4 Specific Outputs from this Study

1.4.1 The Commission has requested the following specific outputs from this study:

- i) On the basis of existing facts and figures relating to the number of cultural heritage items to be digitised, provide an estimate of the total cost of Digitising Europe's cultural heritage and relate the findings to existing resources for Digitisation.
- ii) Provide estimations for average Digitisation costs for cultural heritage assets of different formats. This estimation should be based on real costs in relation to specific formats and quality requirements demanded by Europeana. The estimations should be presented in a form to allow cost estimation for varying numbers of items and formats such as books, paintings, audio files, video files etc. Furthermore the costs for digitising specific quantities (eg. 100,000 monographs) should also be compared to other costs for public investments in the field of culture or elsewhere.

1.4.2 In the development of the outputs, we have produced a Digitisation Costs Calculator, which is based on the figures referred to throughout this report.

1.5 What this Study is not

1.5.1 Our aim in producing this study is not to break new ground, nor to undertake significant new research into the costs of Digitisation. Our figures are based on existing published sources, and all of the tolerances and margins for error associated with those sources affect this material equally.

1.5.2 It is clear from the outset that there is no single authoritative answer to the question of how much it would cost to digitise Europe's cultural heritage. Digitisation is a process and as with all processes, the actual costs depend on a wide variety of variables. We have created broad estimated costs based on the published data available, and applied these to a set of potential models to create illustrative scenarios. The figures here should not be used as an absolute guide to costing a specific project.

1.5.3 What we have created here is a set of estimated costs based on an extrapolation of figures over a 10-year period of funded Digitisation in Europe. It is not, nor is it intended to be, statistically definitive. It is an estimate created for the purpose of supporting the development of new policy and programmes within the European Commission.

- 1.5.4 Critically, in this report, we have only addressed the parts of the Digitisation Lifecycle that relate to the selection, preparation, description and scanning/photography of material, and the creation of surrogates (such as PDF and OCR files) which enable access to it. The figures here do not take account of the long-term or lifetime costs of owning, managing, distributing and preserving it.
- 1.5.5 Anyone investing in Digitisation should take into account not only the costs of creation, but also the total lifetime cost of ownership, which is widely estimated to be as much as 50-100% of the costs of the original digitisation (falling to 10-25% in the case of academic institutions with existing preservation infrastructure).
- 1.5.6 To fail to account for this lifetime cost would be to fail to acknowledge the duty of care to the newly-created digital assets and would lead, ultimately, to a generation of unsustainable or lost material.

2. Defining Digitisation

2.1 What is 'Digitisation'?

- 2.1.1 Digitisation is a loosely-defined term which describes the set of management and technical processes and activities by which material is selected, processed, converted from analogue to digital format, described, stored, preserved and distributed.
- 2.1.2 In this sense, Digitisation is an example of a *supply-chain* activity – one which generates an output (a product) based on a managed input (raw materials) which is distributed and transacted with an end-user.
- 2.1.3 In the European cultural heritage sector, Digitisation has come to signify the various activities through which physical (analogue) cultural content, such as books, artefacts, records and other cultural material are translated into a digital form, described and made accessible through digital channels such as the Internet.
- 2.1.4 It is here that the first of a number of critical distinctions is necessary. Although it refers to similar processes in museums, archives and libraries, the basic conceptual model differs in each domain. Hence while Digitisation of published material in libraries is best characterised as a form of *conversion* or *replacement* (converting essentially the same material and content from one display/storage format to another without a significant loss of its cultural value, meaning or significance), in the case of museums and archives it is better characterised as *surrogacy* (the creation of a digital image and metadata which records and represents the original object, record or document).
- 2.1.5 The distinction is important because conversion is mass-scale, reproducible and broadly *lossless* in the sense that the information content of the original is not fundamentally lost in the process. Surrogacy, on the other hand, is less amenable to mass-production workflows because of the inherent complexity and variability of the material.
- 2.1.6 It is, however, important not to regard too strictly the distinction between Digitisation in museums, archives and libraries (particularly in the case of National institutions), given that the nature of the material they are managing is frequently the same.
- 2.1.7 For the purposes of this study, we will apply the Supply Chain Model, in order to identify the interdependent processes, systems and infrastructures for Digitisation, which we will use as the basis of an activity-based cost model.

2.2 Inputs and Outputs

- 2.2.1 As with any manufacturing or supply process, the total costs will vary significantly according to three critical factors:
- The nature, complexity and fragility of the input (the material to be digitised)
 - The operational efficiency (and repeatability) of the management & production processes
 - The quality, scope, complexity and durability of the output (the digital assets and related metadata)
- 2.2.2 The past decade has seen very significant changes across all three of these variables. On the input side, as cultural heritage organisations have continued to acquire new material, and increasingly born-digital material, the nature of the holdings of cultural heritage institutions has fundamentally changed.
- 2.2.3 In terms of the management and production processing of digital content, there have been very significant changes across the sector. Over many years, the quality, efficiency and availability of basic Digitisation infrastructure (equipment, skills, space) have improved considerably, which in turn is allowing the cultural heritage sector to benefit from economies of scale and repeatability (the economies of mass-production).

- 2.2.4 The range of potential uses for digital cultural content, and hence the relative complexity of choosing appropriate output formats has also increased exponentially in the past decade, and looks set to increase. This is mostly due to changing circumstances which the cultural heritage sector itself cannot influence, such as changes arising through research, development and innovation in the ICT industry and wider content markets.
- 2.2.5 For the purposes of this report, we have taken as our baseline the European Commission's requirement to assess costs of Digitisation to the *quality standards required by Europeana*. Given that Europeana is essentially an aggregator of descriptive metadata, these quality standards are relatively low – they do not require, for example, that every asset recorded in Europeana should correspond to a preservation-quality archival reference image.
- 2.2.6 Here, it should be recognised that there is an important distinction between orthodox, canonical Digitisation to support preservation, curation and research and *digital content creation*, which is concerned with the creation of media assets and information that can be used for discovery, re-use, creativity and entertainment.
- 2.2.7 The traditional view is that the proper output of Cultural Heritage Digitisation ought to be rich, preservation-quality assets that can subsequently be resolved down to create different output formats. While this is certainly optimal in terms of maximising the sustainability and re-purposability of the material, it is less so in terms of cost (the implication being that all Digitisation should be to the highest standards of quality, irrespective of the likelihood of the material being used).
- 2.2.8 A key consideration, therefore, should be on whether the proposed cost models should relate to canonical Digitisation or '*good enough*' Digitisation to provide a useful output. Considering the scale of the challenge, it may be that the cultural heritage sector could consider a sliding scale of prioritisation (as they do already with the storage of their physical collections) whereby the smaller 'elite' of culturally significant material is Digitised to a higher standard than the long-tail of less important material.

2.3 The Importance of Context

- 2.3.1 Cost is not solely a function of input, process and output. It also changes over time and dependent on the strategic, organisational and professional context within which it happens.
- 2.3.2 Hence the cost of Digitisation in the context of a task-and-finish project (in which all of the startup, capital, production and distribution costs have to be allocated within the project budget) is very different from the cost of Digitisation within an organisation that has made a long-term strategic investment in and commitment to Digitisation as a core function.
- 2.3.3 A significant factor, therefore, when modelling costs, rests in the existing sunk costs of investment in previous Digitisation activities and the *unseen* variable of an organisational context which includes, for example, technical competence, pre-existing relationships between teams, management experience and knowledge of the legal framework around copyright.
- 2.3.4 In other words, Digitisation costs should not be assessed simply *per se*, but they should be adjusted to reflect the capacities and experience of the company or organisation that is responsible for the Digitisation.
- 2.3.5 It is also worth considering, from a European perspective, whether long-term strategic investment in Digitisation infrastructure, capacity and skills (for example, through the Competence Centres already proposed by the Commission) would not deliver a greater return on the Commission's investment than individual Digitisation projects, given the intangible but significant economic advantage accruing over time.

2.4 Quality

- 2.4.1 Quality is a critical and often overlooked variable which has a profound effect on the eventual costs of a Digitisation project or programme.
- 2.4.2 Quality itself is partly a function of the expected use of the digital material, but it also exerts an influence throughout the digital content supply-chain. An individual scanner operator's expectations of quality, for example, will have an influence on their workflow rate and efficiency.
- 2.4.3 The quality and integrity of born-digital material will have an influence on how rapidly it can move through the production and descriptive process. On a very basic level, the quality of the documentation of the source material has a profound impact on the efficiency of selection and description during pre-production processes.
- 2.4.4 Quality also exerts a very fundamental influence on *what is digitised*. Most cultural institutions prioritise Digitisation activities either thematically or according to an internal calculus of the significance and importance of the material.
- 2.4.5 Hence Digitisation itself is proving to act as a form of *natural selection* for cultural material, in which decisions are being made about which of an institution's holdings it is most immediately pressing to record or convert into a digital form. Given the increasing influence of the Web on society, it is likely that material not digitised during this period is effectively marked as being of relatively little cultural value (in the same way as material in store is at risk of being prioritised for disposal). This places a burden of responsibility on cultural heritage institutions who, in selecting whether a given book, record or artefact is to be subject to Digitisation is effectively deciding whether information about that material is to be available to users in the future.
- 2.4.6 Again, it is important to point out that this calculus of cultural value and significance operates very differently in different industry sectors. Hence while it might be legitimate for a museum to prioritise the Digitisation of a smaller number of masterpieces, the value proposition for a library (or a project such as *Google Books*) will be to achieve the conversion of as many individual books as possible without as great a distinction around quality or cultural significance.
- 2.4.7 It is an interesting side-note to appreciate that Europeana represents an attempt to reconcile both sets of priorities – achieving breadth of coverage while also seeking to ensure depth of representation through access to masterpieces and other culturally significant works.

3. Cost-Models for Digitisation

3.1 Different Approaches

- 3.1.1 A number of European and US projects have developed sophisticated cost models for Digitisation, based on the extrapolation of actual and inferred cost data from a variety of Digitisation projects.
- 3.1.2 Most of these models are based on various forms of *activity-based costing*, that is, they analyse the processes of Digitisation in terms of their constituent activities (see below), and assign costs to each based on a number of variables and adjustments to allow for, for example, the scale, complexity and format of the material to be Digitised.
- 3.1.3 In addressing the central question of this study, we will make use of the findings and calculations of several of these methodologies, including:
- DiCoMo¹
 - NUMERIC²
 - JISC *Digitisation Costs Study*³
- 3.1.4 In addition, we will make use of cost and pricing models from commercial Digitisation services, as well as cost data (where available) from commercially-funded or public/private partnership projects.

3.2 Digitisation Activities

- 3.2.1 As with any other manufacturing or supply process, the process of Digitisation can expand or contract to include a variety of activities, depending on the nature of the activity.
- 3.2.2 In order to create a meaningful and consistent approach to this enquiry, the authors have defined the scope of Digitisation as including the following activities⁴:

<i>Activity</i>	<i>Description</i>
Selection	Choosing material to be digitised
Preparation	Making objects and books ready to be digitised
Description	Cataloguing, description, indexing and the creation of management information
Conservation	Care, handling, packaging, transport and conservation of the material
Production of Intermediates	For example, microfilming and photography
Technological Infrastructure	Includes equipment (scanners, computers), software and suitable space for Digitisation

¹ Digitisation Cost Model (DiCoMo) - <http://cat.inist.fr/?aModele=afficheN&cpsidt=17182074>

² NUMERIC Digitisation Costs study - <http://www.numeric.ws/>

³ JISC *Digitisation Costs Study* - <http://www.jisc.ac.uk/media/documents/programmes/digitisation/digitisation-costs-full.pdf>

⁴ MINERVA *Digitisation Cost Reduction Handbook* -

http://www.minervaeurope.org/publications/CostReductioninDigitisation_v1_0610.pdf

Quality Management	Error checking and correction
Conversion to master Digital formats	Scanning, digital photography or audio and video encoding
Storage/maintenance	Storage and management of digital assets for use and preservation

3.2.3 As can be seen from the above schedule of costs, the *types* of costs involved in Digitisation can be grouped as:

- Creation/conversion costs:
 - Overhead (staffing, space, depreciation on equipment, quality control)
 - Fixed capital expenditure (equipment, training, software licensing)
 - Variable production costs (per-item scanning, rights clearance)

- Long-term Management costs:
 - Overhead (staffing, space, depreciation on equipment)
 - Capital expenditure (equipment, storage infrastructure, training, software)
 - Variable retro-conversion costs (format-shifting, management)

3.2.4 The LIFE Project has provided a useful model of the nature of the costs incurred in Digitisation and Digital Curation:

Acquisition	Ingest	Bit-stream Preservation	Content Preservation	Access
Selection	Quality Assurance	Repository Admin	Preservation Watch	Access Provision
Submission Agreement	Metadata	Storage Provision	Preservation Planning	Access Control
IPR & Licensing	Deposit	Refreshment	Preservation Action	User Support
Ordering/invoicing	Holdings Update	Backup	Re-ingest	
Obtaining	Reference Linking	Inspection	Disposal	
Check-in				

LIFE Model v2.1(<http://www.life.ac.uk/3/docs/ipres2009v24.pdf>)

- 3.2.5 For the purposes of this enquiry, we have focussed on the primary creation/conversion costs arising from Digitisation. It is important, however, to note that funding these upfront fixed and variable costs of *creating* digital content in turns creates a significantly greater financial obligation to the long-term preservation, management and publication costs of this material.
- 3.2.6 In this sense, funding models for Digitisation are directly analogous to those of physical museum, archive and library infrastructure. Funding acquisitions without provision for the lifetime costs of ownership and stewardship of the material has a negative long-term effect on the integrity of the cultural record. In the same way, funding the creation of digital content without a clear economic model for supporting the long-term stewardship costs will result in the long-term loss of the value of the initial investment.
- 3.2.7 Hence, as a corollary to the question ‘how much would it cost to digitise the cultural heritage of Europe?’ is a further question (not addressed in this study) of ‘how much would it cost to ensure that the digitised record of Europe’s cultural heritage will still be available in 10 years time?’. The response to this second question depends on the extent to which cultural heritage organisations internalise the responsibility to curate and conserve digital material alongside their existing collecting practices.

4. Understanding Formats

4.1 The Impact of Format & Condition on Cost

- 4.1.1 The scope of material collected, managed and curated by cultural heritage institutions spans the diversity of human creativity and output. Hence, the nature of the potential input materials for Digitisation is tremendously diverse.
- 4.1.2 Format and condition have a direct and significant impact on unit costs. Where material adopts a broadly standardised format (such as printed books, for example), it is possible to achieve cost reductions through the use of repeatable, large-scale, batch or semi-industrial processes. Where the material displays a wide range of formats, or makes use of a format that precludes batch-processing (such as the need to light natural specimens individually, for example) then the costs will be significantly higher.
- 4.1.3 Most Digitisation workflows include an element of pre-scanning preparation, conditional assessment and potentially also conservation. Where objects have to be moved to be digitised (as is the case in most projects), and particularly where the Digitisation is taking place offsite, additional costs of inspection, packaging, transport and handling must be taken into account.
- 4.1.4 Not only this, but format has a direct impact on the extent of the material that is to be considered for Digitisation. Hence, an estimated 30% of AV material in collections will show some signs of deterioration and up to 25% is considered too fragile to be exposed to air and light in the Digitisation process.
- 4.1.5 A simple classification scheme for formats has been developed by a number of cultural heritage agencies to include:

<i>Classification</i>	<i>Scope</i>
Books (simple)	Simple, unbound or bound books of a standard size, mainly text in a standard sans-serif font (eg. paperback fiction books)
Books (complex)	Non-standard books with complex binding, images or typography (eg. reference books, dictionaries, non-fiction hardback)
Simple 2D	Simple flat documents on a standard paper stock with relatively simple text content
Complex 2D	Non-standard sizes or media or with relatively complex graphical or schematic content (eg. manuscripts)
Simple 3D	Simple 3D objects with relatively few planes/facets (eg. biscuit tins, coins)
Complex 3D	Complex 3D objects with numerous planes/facets/indentations and/or structural or operational elements (eg. medical or scientific instruments)
Large-scale 3D	Larger and/or working 3D objects with complex spatial requirements (eg. vehicles)
Simple Audio	Simple, well-recorded audio content stored on a stable medium
Complex Audio	Complex, multi-part or poorly-described audio stored on a fragile medium
Simple Video	Simple, well-recorded video content stored on a stable medium
Complex Video	Complex, multi-part or poorly-described video stored on a fragile or medium

- 4.1.6 It should be noted that there is no single authoritative system of classification for formats, and the approach tends to vary considerably between sectors. Hence, 'complex 2D', which might include both maps and manuscripts, are often separated into its constituent parts.
- 4.1.7 For the purpose of this study, we will adopt the classification used in the EU-funded NUMERIC project, since this has been tested most robustly against actual Collection Types, although for the purposes of clarity, we will disaggregate 'rare' books from in-print and out-of-print books in current circulation.
- 4.1.8 The NUMERIC classification is shown in the table below:

Object/items/materials in collections	Units
Books (incl. 'Rare Books')	Published, printed books
Newspapers	Published, printed newspapers
Journals and other serials	Periodicals, magazines or other published material in a serialised form
Government publications	Printed material produced by Government or Government agencies
Other printed material	Other printed material not confirming to any of the above definitions
Manuscripts	Handwritten book or document
Maps	Map of geographical areas or features in a printed form
Photographs	Photographic reproductions, produced in hard copy
Microforms	Printed micro-reproductions of works such as newspapers
Engravings	A print made from an engraved plate
Drawings	An illustration drawn by hand in a hard copy form
Posters	Printed posters
Postcards	Printed postcard
Sheet music	A musical composition in printed or written form
Other images not listed above	Other forms of visual representation not included in the above list
Archived Government records	Information records produced by Government or Government agencies
Archived historic records	Information records of particular historical significance
All other archived records	And other form of information record not included above
Man-made artefacts in museums	Man-made objects, tools, artefacts included in museum collections
Natural specimens	Natural material (flora, fauna, mineral) in museum collections
2D works of art	Works of art in a 2D format, such as paintings, illustrations & prints
3D works of art	Works of art in a 3D format, such as sculptures or other fabrications
Other museum objects	Other objects in collections of a type not confirming to the above list

Film and video recordings	Visual material recorded onto a physical storage medium
Music and recorded sound	Audio material recorded onto a physical storage medium
Other items not listed above	All other material not included in the above list

4.1.9 The table below shows the NUMERIC formats mapped to our 'simplified' formats list:

<i>Simple Classification</i>	<i>NUMERIC Classifications</i>
Books (simple)	<ul style="list-style-type: none"> • Books
Books (complex)	<ul style="list-style-type: none"> • Rare Books
Simple 2D	<ul style="list-style-type: none"> • Journals • Posters • Postcards • Archived Government records • Other printed material • Other archived material
Complex 2D	<ul style="list-style-type: none"> • Newspapers • Manuscripts • Maps • Photographs • Engravings • Drawings • Sheet Music • Microforms • 2D works of Art • Archived historic records
Simple 3D	<ul style="list-style-type: none"> • Man-made artefacts (simple) • Other museum objects
Complex 3D	<ul style="list-style-type: none"> • Man-made artefacts (complex/fragile) • Natural history specimen • 3D works of Art • Monuments • Landscape
Large-scale 3D	<ul style="list-style-type: none"> • Man-made artefacts (large) • Natural history specimen (large)
Simple Audio	<ul style="list-style-type: none"> • Audio recordings (simple)
Complex Audio	<ul style="list-style-type: none"> • Audio recordings (multi-part, fragile)
Simple Video	<ul style="list-style-type: none"> • Video recordings (simple) • Film footage • Other time-based visual media
Complex Video	<ul style="list-style-type: none"> • Video recordings (complex, multi-part, fragile)

5. A Structured Approach to Normalising Costs

5.1 Developing a Digitisation Costs Calculator

- 5.1.1 For the purposes of this study, we have created a calculator of Digitisation costs which enables cultural heritage agencies to input data against a series of weighted variables and multipliers. The structure of this calculator, and the factors of which it takes account, are shown in the table overleaf.
- 5.1.2 It should be noted that none of the figures used in this study are absolute – they are all relative within margins of error, based on the number of variables at play and the relative complexity of their interactions on one another.
- 5.1.3 It is technically not possible to produce a single absolutely authoritative algorithm or formula for the calculation of Digitisation costs given the interdependence of the operational variables at play. Every project is different, and the contextual environment has a tremendous influence on the actual costs.

Table. Schedule of cost variables for Digitisation based on an analysis of actual costs

<i>Group</i>	<i>Variable</i>	<i>Definition</i>
Infrastructure	Organisation type	Selection from a defined list
	Organisation status	Multiple selection from a defined list indicating the legal/governance status of the institution.
	Annual turnover	Annual operating budget for the institution
	Dedicated Digitisation space	Availability of dedicated space for Digitisation
	Dedicated staff	Availability of dedicated staff for Digitisation, including staff attached to partner organisations
	Est. % Collection already Digitised	Estimation of the proportion of Collections already digitised (as an indicator of experience/expertise)
	Dedicated legal support	Availability of dedicated legal advice and support (including capacity and experience in Rights Clearance)
Content	Books (simple)	As described above
	Books (complex)	
	Simple 2D	
	Complex 2D	
	Simple 3D	
	Complex 3D	
	Large-scale 3D	
	Simple Audio	
	Complex Audio	
	Simple Video	
	Complex Video	

Rights Status	Copyright	% of material for which copyright clearances required
	Orphan works	Proportion of material for which the copyright owner is either unknown or cannot be traced
	Out of print	Proportion of material that is known to be out-of-print
Outputs	Hi-resolution image	Whether the Digitisation will produce hi-resolution, archival quality images suitable for preservation purposes
	Promotional image	Whether the Digitisation will produce attractive, specifically-lit images for use in promotional activity
	Lo-resolution image	Whether the Digitisation will produce lo-resolution, archival quality images suitable for online distribution
	OCR output	Whether the Digitisation workflow will use Optical Character Recognition (OCR) scanning
	Encoded video	Whether Digitisation will produce encoded audio in an industry-standard file format
	Encoded audio	Whether Digitisation will produce encoded video
	TEI XML	Whether the Digitisation will produce text with structured markup in Text Encoding Initiative (TEI) XML format
Post-production	Colour correction/balancing	Whether colour correction will be done using photo management software in post-production
	Quality Assurance	Whether resources are allocated to quality management
	Metadata creation	Whether the Digitisation will produce new metadata records about the content
	Metadata enrichment	Whether the Digitisation workflow will enrich existing metadata records about the content

5.1.4 As can be seen from this table, the number of variables involved in costing a Digitisation project is significant, illustrated in the following simple illustration of the options for formats:

<i>Item</i>	<i>Number of potential options</i>
Bound/unbound	2
Page size (8.5x11, 11x17, 17x22)	3
Scanning resolution (300dpi, 400dpi, 600dpi)	3
Scanning bit depth (1, 8, 24)	3
Handling (fragile/non-fragile)	2
Place of performance (on/off site)	2
Possible price combinations	216

Table: Illustrative 'menu' of Digitisation options. (Dan Pence, Systems Integration Group Inc).

5.1.5 In order to provide a meaningful estimated figure for the costs of digitising Europe’s cultural heritage to a quality sufficient that it can be shared with Europeana, we will construct a putative pan-European Digitisation Project with the following specifications:

- That our ‘pan-European’ collection will include material of all formats/types
- That it will span a range of experienced and inexperienced partner organisations
- That it will span a range of international, national, regional and local institutions
- That the desired output of museum and archive Digitisation will be hi-res images and metadata
- That the desired output of library digitisation will be a combination of source images and OCR text in a structured XML format
- That we only intend to digitise one copy of each published title, irrespective of multiples or editions

5.1.6 Given the tremendous variation in cost between in-house, project-based Digitisation and outsourced, ongoing Digitisation (for example through a public/private partnership), we will evaluate the costs of each of these approaches separately. In so doing, we will provide a relative comparison of the cost of funding Digitisation by cultural heritage organisations against the cost of enabling them to outsource or collaborate with commercial partners.

5.2 The NUMERIC Cost Model

5.2.1 The EU-funded NUMERIC project developed a set of unit costs for digitisation (see figure below):

Object/items/materials in collections	Units	Median results			
		Digitised as % of collection	Pages per unit	Cost per unit (€)	Cost per page (€)
Books (incl. ‘Rare Books’)	Volumes	0.05	250	191	0.50
Newspapers	Volumes	1.7	14	-	1.56
Journals and other serials	Volumes	0.1	350	15	0.35
Government publications	Volumes	16	723	2961	3.72
Other printed material	Number	5.9	3	14	3.70
Manuscripts	Number	0.3	123	184	8.74
Maps	Number	1.2	-	11	-
Photographs	Number	3.9	-	4.07	-
Engravings	Number	4.4	-	20	-
Drawings	Number	20	-	4.82	-
Posters	Number	18	-	17	-
Postcards	Number	33	-	3.14	-
Sheet music	Number	1.3	23	29	1.07
Other images not listed above	Number	18	-	5.13	-
Archived Government records	Metres	50	768	-	0.74
Archived historic records	Metres	25	500	399	0.80
All other archived records	Metres	54	3636	2901	0.80
Man-made artefacts in museums	Artefacts	53	-	3.89	-
Natural specimens	Objects	-	-	-	-
2D works of art	Exhibits	72	-	11	-
3D works of art	Exhibits	75	-	11	-
Other museum objects	Objects	10	-	-	-
Film and video recordings	Hours	3.4	-	1125	-
Music and recorded sound	Hours	9.1	-	14	-
Other items not listed above	Number	45	-	-	-

6. The Scale of European Cultural Heritage

6.1 How much cultural heritage is there?

6.1.1 Estimating the cost of digitising Europe's cultural heritage depends on establishing reasonable estimates to answer the following questions:

- How many museums, libraries and archives are there in the EU*?
- How much material of each type do they hold?
- What is the unit cost of digitising each type of material?
- Based on these figures, what is the total cost of digitising cultural heritage?

* The calculations in this report refer to the specific current membership of the, except where explicitly stated otherwise.

6.1.2 There is no simple way to estimate the quantity of cultural heritage held in Europe's museums, archives and libraries. Efforts towards mapping have been far from comprehensive, and the estimate is complicated by the relative complexity of defining the scope.

6.1.3 In order to reach reasonable estimates, we have broken the domain down into 3 sets of investigations:

Libraries:

- What is the extent and composition of the Library sector in the EU?
- What evidence do we have to quantify the extent of the collections held by different types of library?
- What proportion of Library collections should be/can be digitised?
- What proportion of that body of material has been digitised already?

Archives:

- What is the extent and composition of the Archive sector in the EU?
- What evidence do we have of the extent of the collections held by different types of Archive?
- What proportion of Archive collections should be/can be digitised?
- What proportion of that body of material has been digitised already?

Museums:

- What is the extent and composition of the Museum sector in the EU?
- What evidence do we have of the extent of the collections held by different types of museum?
- What proportion of museum collections should be/can be digitised?
- What proportion of that body of material has been digitised already?

Audio Visual Collections

- What evidence do we have of the extent of the collections of AV material already held in museums, archives and libraries?
- What information do we have about the cost-per-hour of Digitising AV material under differing conditions?
- What are the specific factors affecting Digitisation costs for AV material?
- To what extent can we extrapolate estimated costings for this material in the EU?

6.1.4 Quantifying the amount of material under consideration is not a simple function of the number of each type of institution in a given Member State. There is, for example, considerable duplication of book stock across Libraries, and variation in museums between 100 objects for some leading Art museums up to 110m objects for large Natural History museums.

6.1.5 In addition, there is a significant body of cultural material (for example, film or AV archives) held outside museums, archives and libraries. A comprehensive view on the costs of digitising all cultural material should therefore take some account of the implications of Digitisation by Broadcasters, publishers and other forms of film archive (such as Film Institutes).

6.2 Material not to be digitised

6.2.1 Previous studies have highlighted the fact that there are significant bodies of material in libraries, archives and museums that it is either not appropriate to digitise, or which do not need to be digitised.

6.2.2 This material represents an important adjustment to our calculation of the potential costs of digitising Europe’s cultural heritage, since we can discount from it material that has been identified as not needing to be digitised.

6.2.3 The table below summarises the proportions of Collections that do not need to be digitised, based on the figures established by the NUMERIC project, along with a rough estimate of the proportions remaining to be digitised:

<i>Institution</i>	<i>No need to digitise</i>	<i>Digitisation completed</i>	<i>Awaiting digitisation</i>
Archives	36%	1%	63%
Broadcasters	28%	6%	66%
Museums	3%	25%	72%
Libraries	69%	1%	30%
Other	0%	15%	85%
TOTAL	31%	11%	58%

7. The Cost of Digitising Libraries

7.1 Defining Scope

7.1.1 Our first enquiry will look at the aggregated cost of digitising the material held by European libraries. The first challenge in considering Digitisation in libraries relates to the definition of scope.

7.1.2 European libraries include the following types of organisation:

- National libraries
- Public Libraries
- University Libraries
- School & Special Libraries

7.1.3 There are 45 National Libraries across Europe, as represented by the Council of Europe (totalling approximately 178 service points)⁵.

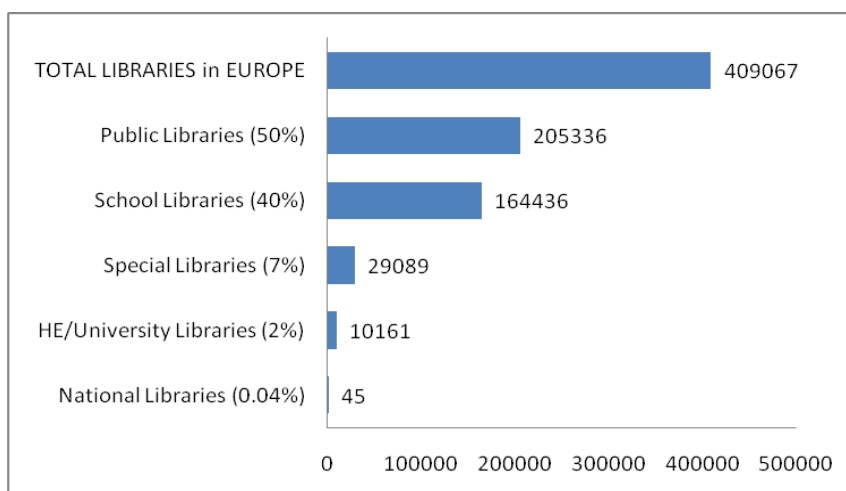
7.1.4 There are approximately 205336 Public Libraries (+/- 7%), at an average of 0.00042 per capita of population.

7.1.5 There are approximately 10161 Higher Education or University Libraries (+/- 7%), at an average of 0.00002 per capita of population.

7.1.6 There are approximately 164,436 School Libraries in the Member States (+/- 7%), at an average of 0.0034 per capita of population.

7.1.7 There are approximately 29089 Special Libraries (excluding Health) in the Member States (+/- 7%), at an average of 0.00006 per capita of population⁶.

7.1.8 The relative distribution of libraries in the EU is shown below:



* Note that 'TOTAL LIBRARIES in EUROPE' also includes service points

⁵ Source: LIBECON/UNESCO Statistics

⁶ Note: Data on Special Libraries is extrapolated based on a single source (1998, LISU/LIST data for UK) and subject to a larger margin of error

7.2 National Libraries

7.2.1 A national library may include a number of additional sites or service points, or in some cases may even be a cluster of smaller institutions. There are an estimated 45 National Libraries in the Council of Europe, representing a total of approximately 178 service points.

7.2.2 National libraries hold large quantities of printed and digital material across a wide spectrum of different formats, including:

- Conference proceedings
- e-resources
- Ephemera
- Incunabula (booklet or pamphlet printed before 1501)
- Journals
- Manuscripts and archives
- Maps
- Microforms
- Moving image
- Music
- Newspapers and comics
- Official publications
- Patents
- Philatelic (stamps)
- Photographs
- Printed books
- Prints and drawings
- Rare books
- Reference works
- Reports
- Sound
- Theses
- Trademarks and designs

7.2.3 National libraries are not normally lending libraries. Instead they collect and care for printed material of particular significance, as well as holding material submitted under requirements for *Legal Deposit* (see below).

7.2.4 Given that many Legal Deposit Libraries also mandate the supply of multiple copies of each new title within its collecting scope, and that Legal Deposit activity is duplicated across multiple institutions, we will assume only one Legal Deposit collection per country, and that this may contain multiple copies of individual titles.

7.2.5 In calculating the likely costs of digitising the holdings of National Libraries, it will be important to consider that most of these institutions have been running active Mass Digitisation Programmes for 5-10 years, and that a significant proportion of their collections are already digitised.

7.3 Public Libraries

7.3.1 The estimated 205,000 Public Libraries throughout Europe fulfil a variety of functions, including lending of printed books, support for learning, provision of Internet access and the provision of access to newspapers, magazines and journals.

7.3.2 For the purposes of calculating the total holdings of books in European libraries (see below), we acknowledge that the primary function of Public Libraries is the provision of access to multiple copies of books, one or more copies of which will be lodged with the Legal Deposit and/or National Library of each Member State.

7.3.3 Hence, when considering the question of Digitising the total number of available titles (as opposed to the total holdings of books *per se*), then we will need to exclude from the calculation the book stock held by Public Libraries.

7.4 University Libraries

7.4.1 As with Public Libraries, when considering the total costs of digitising the collections of Academic/University libraries, it is important to draw a distinction between book stock that is likely to be held already in a Legal Deposit or National Library, published academic literature such as theses or rare books and material that is classed as 'grey' literature (broadly, printed but unpublished material).

7.4.2 Again, for the purposes of this calculation, we will discount the collections most likely to be duplicated elsewhere and focus on the likely costs of Digitisation of the other forms of collections held by University Libraries.

7.5 Special Libraries

7.5.1 There is a large and active community of Special Libraries in Europe, including Company, School and Health Libraries.

7.5.2 School libraries in particular are unlikely to hold significant quantities of material that is not already lodged with a Legal Deposit Library, and so will not be included in this calculation.

7.6 How Many Books?

7.6.1 Since books represent one of the most important (if not actually the most numerous) collection types in Libraries, it is important to establish a reasonable figure for the total number of printed books that require digitisation.

7.6.2 Estimates of the number of actual *books* (as opposed to titles) in Libraries, including editions, duplicates, serials and other multiples vary considerably.

7.6.3 IFLA estimates in 2000⁷, for example, put the total holdings of European Libraries at approximately 16bn items (see table below), including 5.64bn books, at an estimated 1.41 trillion pages.

Media	Quantity (estimated)
Books	5.64 billion books, 1.41 trillion pages
Microform	0.02 billion units
AV	Up to 6.54 billion units
Other	Up to 3.69 billion units
TOTALS	15.89 billion units (+/- 15%)

7.6.4 Given, however, that each library is likely to hold multiple copies of a number of titles, and that ownership of individual titles will be duplicated across multiple institutions, a calculus based on raw book stock would provide a considerable over-estimation of the likely costs of Digitisation.

⁷ IFLA *Global Library Statistics 1990-2000* - http://archive.ifla.org/III/wsis/wsis-stats4pub_v.pdf

- 7.6.5 The Google Books Project has recently completed its calculation of the number of possible titles for Digitisation, giving an estimated total of 130m 'tomes' (a tome being an 'idealized bound volume' of which there may be many millions of copies, but only one 'source' copy). It is worth considering the Google methodology (described below), particularly given that it is based on metadata derived from existing Digitisation programmes.
- 7.6.6 It is also worth considering whether the aim of our idealised pan-European Digitisation programme is to digitise all individual titles within Europe, or whether it is to digitise the material *not already digitised by Google*. Given that we presume that the ultimate aim of the Digitisation is to preserve in perpetuity and to provide free open access to the books, we will not take into account the significant body of material (up to 15m titles to date) that has already been digitised under the Google Books Project.
- 7.6.7 The aim of the Google Books Project is to digitise every book in existence and to make them available through the Google Books platform. In order to coordinate this process, Google undertook a project to estimate the number of books eligible for Digitisation.
- 7.6.8 Google has evaluated the validity of using *International Standard Book Numbers* (ISBN), and their pre-cursors *Standard Book Numbers* (SBN) but discounted these on the basis of inconsistent application and their relatively short history. Similarly, unique identifiers such as *Library of Congress Control Numbers* has also been discounted on the basis that they correlate to bibliographic records, but not to individual titles.
- 7.6.9 Instead, Google aggregates book metadata from 150+ providers (such as union catalogues) to create a raw dataset of a billion records. This is then analyzed (the precise method of analysis is not known) to filter out duplicates within each data provider, giving a total number of 600m de-duplicated records.
- 7.6.10 This 600m records is de-duplicated further using a tiered comparison of different record attributes. Google admits that this is an inexact process, but based on their algorithm, the 600m dataset is reduced further to 210m, which still includes a significant body of records for non-book materials such as maps, microforms, audio recordings and videos. Excluding these results in an estimated 146m 'tomes', which, when filtered to exclude serials, equals approximately 130m unique titles.
- 7.6.11 So how valid is the Google estimate? The logic is clearly based on direct experience of handling actual material, and on the analysis of a significantly greater aggregated body of digital data than exists anywhere else. The formula for the de-duplication of records is as comprehensive as possible given the nature of the data, and from published information is graded according to the statistical reliability of given attributes as comparators.
- 7.6.12 The difficulty is that the calculation ultimately depends on the validity of the input data. Drawing data from 150+ sources lends the calculation critical mass, but there doesn't appear to be any way of independently verifying the correlation between the billion-odd metadata records and actual publishing.
- 7.6.13 The conclusion, then, is that the Google estimate cannot be 100% accurate, but that it is very likely to be the best estimate that anyone is able to provide without a separate Digitisation and metadata-aggregation programme on a comparable scale. It is also worth noting that since the estimate was published in August 2010, very few serious concerns have been raised about its validity.
- 7.6.14 For the purposes of this study, therefore, we will take the Google Books estimate as a comparator against which to consider the validity of the reported data from the libraries themselves, and we will also consider their methodology when considering the issue of filtering to avoid duplication.

7.6.15 We must also be aware of the fact that the Google Books estimate includes *all titles published worldwide*, and not simply those that have been collected by European libraries. It is possible, for example, that the 130m figure includes significant bodies of material from the US, China and the Middle East, copies of which are not held in European libraries.

7.6.16 Even, therefore, if the estimated 130m books is accurate, it is likely that the number of books eligible for digitisation in European libraries is lower than this.

7.6.17 The data from the NUMERIC report, weighted and scaled to the library community *in toto* (as distinct from the relevant institutions covered under the original survey), estimates a total book stock of 215m in the National Libraries.

7.6.18 If we further apply reductions to account for the amount of material *already digitised* and the quantity *not eligible for Digitisation*, we can perform the following calculation:

Description	% reduction	Outcome
Starting figure	-	215m
Adjustment to exclude multiples (based on Google)	11%	191.35
Already Digitised	1%	189.44m
Not suitable for Digitisation	69%	58.73m
Estimated number of books to be digitised	-	58.73m
Estimated number of pages to be digitised (based on average 250 pages per book)	-	1.47bn pages

7.6.19 This figure is not entirely reliable, since it depends on comparisons between a number of different datasets within which different models for filtering may already have been applied. In particular, the figure of 69% of books not eligible for Digitisation may already include an element of filtering for duplicates. If, on the other hand, we postulate an upper limit of 69% and a lower limit of 50% (reported in NUMERIC), we can produce the following ranges:

Description	Multiplier	Calculation
Starting figure	-	191.35m
Already digitised	1%	189.44
If 69% not eligible for Digitisation	189.44m less 69%	58.73m
If 50% not eligible for Digitisation	189.44m less 50%	94.72m
Upper estimate for total number of pages		2.36 bn pages
Lower estimate for total number of pages		1.47 bn pages

- 7.6.20 Based on these calculations, we can postulate a total *eligible* quantity of individual titles for Digitisation in libraries (discounting duplicates and duplicated stock in Public and University Libraries) at between 59m and 95m (noting that the data available for this calculation is subject to a significant margin of error).
- 7.6.21 This means that the total number of *pages eligible* for Digitisation ranges from 1.47bn up to 2.36bn.
- 7.6.22 Taking the *mean* of these figures provides an estimated 77m individual titles held in EU libraries that are eligible for Digitisation, or 1.93bn individual pages.
- 7.6.23 A final consideration is the fact that as an increasing quantity of published material is born-digital, and an increasing number of countries have arrangements for *Electronic Legal Deposit*, the proportion of new book publishing that does not require Digitisation is increasing. This means that, ultimately, the effort to digitise the total holdings of books is predominantly a *legacy* effort.

7.7 How Much Per Page?

- 7.7.1 Having established a reasonable estimated range for the number of pages eligible for Digitisation within the scope of this study, it remains to establish a reasonable calculation of the estimated costs per page to digitise them.
- 7.7.2 As noted previously, estimates for per-page Digitisation costs vary considerably, both in actual value and in terms of the costs that are allocated under them.
- 7.7.3 It is difficult, for example, to provide a valid like-for-like equivalence between the costs of Digitisation in a project context (where factors such as overheads, startup and capital costs must be apportioned across the per-page cost), an outsourced Digitisation service context (where overheads may be defrayed across a broader client base) and a large-scale Mass Digitisation programme taking place in the context of a Public Private Partnership (in which some overhead costs can be presumed because they are accounted for elsewhere within the private partner enterprise).
- 7.7.4 Hence, in order to establish meaningful multipliers, we propose to investigate the per-page costs for Digitisation in the following 3 models:
- Project-based Digitisation of all eligible, non-duplicated book stock held in European libraries (National and Legal Deposit libraries) in which the Digitisation projects are run on a distributed basis by a large number of individual organisations and Digitisation is performed in-house and onsite.
 - Project-based Digitisation of all eligible, non-duplicated book stock held in European libraries (as above) in which Digitisation is outsourced to a commercial offsite Digitisation service.
 - Long-term mass-Digitisation Programme of all eligible, non-duplicated book stock held in European libraries (as above) run as a Public/Private Partnership with a major technology partner, and in which the majority of Digitisation takes place onsite.
- 7.7.5 These models, and the cost models associated with them are, of necessity, a simplification. They do, however, highlight the relative value-for-money of working in partnership with external services and agencies, in which the overall costs of Digitisation can be defrayed over multiple years and the overheads attributable to the process for the library itself are significantly reduced.

- 7.7.6 One critical point to note is that these estimates do not include the costs of retrospective Rights Clearance for the material to be digitised. The costs of clearance are outside the scope of this report, and have been addressed elsewhere in research on behalf of the European Commission.
- 7.7.7 The output specification for our putative Digitisation project is that the output materials should be of a quality sufficient to satisfy the 'quality specifications for Europeana'. Since, however, Europeana is a metadata aggregator, its requirements in terms of image quality and metadata are significantly lower than, for example, the specifications for preservation formats.
- 7.7.8 For the purposes of this exercise, we will assume that the outputs from the proposed Digitisation programme would include:
- One METS file per book
 - One PDF file per book
 - One ALTO or equivalent file per page (containing the OCR text)
 - One JPEG2000 file per page (for archiving)

7.8 Case 1: Project-based In-house Digitisation of All Eligible Books

7.8.1 Assuming a total of 59-95m books (1.47 to 2.36bn pages) to be digitised, we can calculate the likely costs of Digitisation, based on the following elements:

- Capital costs (equipment, setup, staffing, training)
- Preparation, including unbinding (per volume cost)
- Scanning costs (per page, bitonal) x 250
- OCR conversion costs (per page) x 250
- PDF conversion costs (per page) x 250
- Simple metadata creation (per volume)

7.8.2 The component costs vary considerably, but the following high and low estimated figures are based on averages from current and recent projects:

Low Estimate

<i>Item</i>	<i>Per-volume Cost (€)</i>
Capital costs (per volume)	4.50
Preparation, selection and unbinding (plus conditional assessment)	4.50
Scanning costs (€0.15 per page x 250 pages)	37.5
OCR conversion costs (€0.08 per page x 250 pages)	20
PDF conversion costs (€0.2 per page x 250 pages)	50
Simple metadata creation (per volume, simple DC record)	7.5
TOTAL per volume	€124 per book

High estimate:

Item	Per-volume Cost (€)
Capital costs (per volume)	6
Preparation, selection and unbinding (plus conditional assessment)	4.50
Scanning costs (€0.23 per page x 250 pages)	57.50
OCR conversion costs (€0.08 per page x 250 pages)	20
PDF conversion costs (€0.3 per page x 250 pages)	75
Simple metadata creation (per volume, simple DC record)	7.5
TOTAL per volume	€170.50 per book

7.8.3 Based on these high and low estimated costs, and the upper and lower limit of our estimated number of eligible books, the potential cost ranges are illustrated in the following table:

	Books (lower estimate)	Books (upper estimate)
Cost per book (lower estimate)	€7.32bn	€11.78bn
Cost per book (upper estimate)	€10.06bn	€16.20bn

7.8.4 The total cost range, therefore, for digitising all *eligible* books in libraries in EU Member States is approximately €7.32bn to €16.20bn, on the basis that the work is completed predominantly in-house on a project-basis and requiring upfront capital investment in setup costs.

7.8.5 Given the sensitivity of Digitisation costs to external and organisational factors, a direct mean is less relevant here, but for the sake of completeness, the mean of these figures would be €11.76bn.

7.8.6 It should be noted that these calculations provide a lower estimated figure for the per-volume cost of book Digitisation than the original findings of the NUMERIC report, which provided a median estimate of €191 per volume and a much higher top-end estimate.

7.8.7 One possible reason for the difference in costs is that some elements of the cost model have reduced since the original NUMERIC survey exercise, for example because of reductions in equipment or software costs. Equally, the scope of data gathering for this Digitisation Costs study has been far less than that gathered for NUMERIC. It is to be hoped that the new ENUMERATE network will provide a robust evidence-base with which to refine these cost estimates.

7.9 Case 2: Project-based Outsourced Digitisation of All Eligible Books

7.9.1 Assuming a total of 59-95m books (1.47 to 2.36bn pages) to be digitised, we can calculate the likely costs of Digitisation, based on the following elements:

- Project Management (predominantly staff costs and overhead)

- Preparation, including unbinding (per volume cost)
- Scanning costs (per page, bitonal) x 250
- OCR conversion costs (per page) x 250
- PDF conversion costs (per page) x 250
- Simple metadata creation (per volume)

7.9.2 The key differentiating factors between the costs under this model and the previous model are the comparatively lower upfront capital costs on the acquisition and installation of equipment, recruitment, staffing overheads and training.

7.9.3 The component costs vary considerably, but the following high and low estimated figures are based on averages from current and recent projects:

Low Estimate

<i>Item</i>	<i>Per-volume Cost (€)</i>
Project management (per volume)	2.15
Preparation, selection and unbinding (plus conditional assessment)	4.50
Scanning costs (€0.15 per page x 250 pages)	37.5
OCR conversion costs (€0.03 per page x 250 pages)	7.50
PDF conversion costs (€0.07 per page x 250 pages)	17.5
Simple metadata creation (per volume, simple DC record)	4
TOTAL per volume	€73.15 per book

High estimate:

<i>Item</i>	<i>Per-volume Cost (€)</i>
Project Management (per volume)	3.50
Preparation, selection and unbinding (plus conditional assessment)	6
Scanning costs (€0.23 per page x 250 pages)	57.5
OCR conversion costs (€0.03 per page x 250 pages)	7.50
PDF conversion costs (€0.10 per page x 250 pages)	25
Simple metadata creation (per volume, simple DC record)	4
TOTAL per volume	€103.50 per book

7.9.4 Based on these high and low estimated costs, and the upper and lower limit of our estimated number of eligible books, the potential cost ranges are illustrated in the following table:

	Books (lower estimate)	Books (upper estimate)
Cost per book (lower estimate)	€4.32bn	€6.95bn
Cost per book (upper estimate)	€6.11bn	€9.83bn

7.9.5 The total cost range, therefore, for digitising all *eligible* books in libraries in EU Member States is approximately €4.32bn to €9.83bn, on the basis that the work is completed predominantly via outsourcing and without the additional requirement for upfront capital expenditure on setup costs.

7.9.6 A *mean* across these figures would give an estimated cost of €7.08bn.

7.10 Case 3: Digitisation of All Eligible Books by Public/Private Partnership

7.10.1 Assuming a total of 59-95m books (1.47 to 2.36bn pages) to be digitised, we can calculate the likely costs of Digitisation, based on the following elements:

- Project Management (per volume)
- Preparation, including unbinding (per volume cost)
- Scanning costs (per page, bitonal) x 250
- OCR conversion costs (per page) x 250
- PDF conversion costs (per page) x 250
- Simple metadata creation (per volume)

7.10.2 The key differentiating factors between the costs under this model and the previous two are the comparatively lower per-scan, OCR conversion and metadata creation costs. These are based on figures from the Google Books and Microsoft Digitisation projects, and the cost differential is most likely due to the increased processing rate, slightly lower error rate and the economies arising from experienced staff and sunk investment in technology.

7.10.3 The component costs vary considerably, but the following high and low estimated figures are based on averages from current and recent projects:

Low Estimate

<i>Item</i>	<i>Per-volume Cost (€)</i>
Project Management	2.15
Preparation, selection and unbinding (plus conditional assessment)	4.50
Scanning costs (€0.08 per page x 250 pages)	20
OCR conversion costs (€0.03 per page x 250 pages)	7.50
PDF conversion costs (€0.07 per page x 250 pages)	17.50

Simple metadata creation (per volume, simple DC record)	3.50
TOTAL per volume	€55.15 per book

High estimate:

<i>Item</i>	<i>Per-volume Cost (€)</i>
Project Management	3.50
Preparation, selection and unbinding (plus conditional assessment)	4.50
Scanning costs (€0.1 per page x 250 pages)	25
OCR conversion costs (€0.04 per page x 250 pages)	10
PDF conversion costs (€0.08 per page x 250 pages)	17.50
Simple metadata creation (per volume, simple DC record)	3.50
TOTAL per volume	€66.50 per book

7.10.4 Based on these high and low estimated costs, and the upper and lower limit of our estimated number of eligible books, the potential cost ranges are illustrated in the following table:

	Books (lower estimate)	Books (upper estimate)
Cost per book (lower estimate)	€3.25bn	€5.24bn
Cost per book (upper estimate)	€3.92bn	€6.32bn

7.10.5 The total cost range, therefore, for digitising all *eligible* books in libraries in EU Member States is approximately €3.25bn to €6.32bn, on the basis that the work is completed predominantly via outsourcing and without the additional requirement for upfront capital expenditure on setup costs.

7.10.6 The *mean* of these costs would amount to €4.79bn.

7.11 Summary of Costs

7.11.1 The following table summarises the estimated cost ranges for the 3 different approaches to mass-digitisation, based on the above calculations:

	Low estimate (€bn)	High estimate (€bn)	Mean (€bn)
In-house, project-based Digitisation	7.32	16.20	11.76
Outsourced, project-based Digitisation	4.32	9.83	7.08
Public-Private Partnership Digitisation	3.25	6.32	4.79

- 7.11.2 It is important, however, to appreciate that these costs (apart from the fact that they are based on very broad approximations and not exact figures) are not exactly comparable on a like-for-like basis.
- 7.11.3 The costs of in-house, project-based Digitisation are significantly higher, but this is due in part to the huge upfront capital investment of equipping an institution to undertake Digitisation. This is, however, a one-off capital expenditure that can be defrayed over a number of years. Not only this, but the hidden value of this approach is that it generates core skills and expertise and increases internal capacity to undertake future Digitisation activities. The benefits of this internal development are not simply in enabling more efficient future Digitisation, but also in improving the organisation's own knowledge of its Collections.
- 7.11.4 On the other hand, the apparently much lower costs of Digitisation in the context of a Public Private Partnership masks the extent to which capital expenditure, overheads and 'hidden' costs (such as the costs of acquiring expertise in IPR and licensing) are offset by existing sunk investment on the part of the commercial partner.
- 7.11.5 It is also important to note that the output of the first two types of project (which generate digital assets that are the 'property' of the cultural institution and hence form part of its public collections) is different from the output of the latter, PPP-project, in which the commercial partner may place restrictions on the use of the output (including, for example, commercial use by the cultural institution).
- 7.11.6 Care should be taken, therefore, when considering the optimal approach, not to be too directly influenced by the significant apparent variations in cost. It is of equal importance to consider the infrastructural and capacity benefits to cultural institutions of developing internal Digitisation programmes, as well as the usability of the end-product to satisfy an open-access or Public Good mandate.

7.12 Orphan Works

- 7.12.1 'Orphan' works are works for which the owner of the Copyright is either unknown, or cannot be traced.
- 7.12.2 Orphans come into being for a number of reasons. Either the information about Copyright attribution is not collected at the point of acquisition, or the information becomes disconnected from the material during its lifetime, or it is simply the case that the original rightsholder can no longer be traced in order to secure appropriate permissions.
- 7.12.3 The issue of Orphan Works is of particular importance to cultural institutions because they present a tension between the mandate to provide open public access and the need to secure appropriate permissions for the use of the material.
- 7.12.4 The choice, for many, is either not to digitise these works (and therefore not to make them accessible for search and discovery) or to digitise them and thereby to accept a risk of infringing copyright, and the potential legal liability arising from such an infringement.
- 7.12.5 In many cases, cultural institutions have opted not to digitise material that is of unknown or unverified derivation, and hence the issue is having a direct effect on the completeness and integrity of the cultural record. Several solutions have been discussed, on both a national and a European level, including collective licensing and indemnification. None of these have so far found widespread acceptance, due in part to the lack of legal support for the principles of Due Diligence and Diligent Search.
- 7.12.6 The problem of Orphan Works is less acute in the case of recent printed books than for in-copyright museum collections, for example, but it is nevertheless a significant factor in light of the considerable opportunity cost of not providing access to material of unknown derivation.

7.12.7 The joint Strategic Content Alliance/Collections Trust report conducted the *In from the Cold*⁸ research project to assess the range and impact of the Orphan Works problem on the museum, archive and library sector. The table below shows the averaged-out results of key findings in this research:

	Museum	Library	Archive	Gallery	Education	Health
Mid-range estimate of average number of works that are Orphans	7.5%	7.5%	25.5%	7.5%	7.5%	2.5%
Proportion of institutions whose services are seriously affected by Orphan Works	95%	88%	94%	96%	91%	58%

7.12.8 Based on these estimates (which were confirmed in a recent exercise to extend the findings across European countries other than the UK), we estimate that between 4.43m and 7.13m of the 59m to 95m titles eligible for Digitisation in European libraries may be Orphan Works.

7.12.9 If the IFLA projection of a total 5.64bn books in stock in libraries is correct, this could imply a figure of some 0.4bn titles in public ownership that are Orphan Works, although this figure would need to be subject to far more rigorous investigation before it could be considered valid.

7.12.10 Nor is the Orphan Works issue limited to books. An estimated 95% of newspapers published before 1912 are now orphans, and the *In from the Cold* study estimated approximately 90% of the photographic record as orphaned.

7.12.11 Collectively, then, Orphan Works represent a huge proportion of the holdings of European cultural institutions, and represent a significant barrier to future accessibility.

7.12.12 In the report *Assessment of the Orphan Works issue and the Costs for Rights Clearance*⁹, Anna Vuopala found examples in which the potential costs of Rights Clearance for Orphan Works were 20-50 times higher than the initial costs of Digitisation. Not only this, but the ‘hit rate’ – the rate at which the rightsholder is successfully traced and appropriate permissions secured can be less than 50%, even after costly and diligent search.

7.12.13 As with many issues around Digitisation, it is more likely that an economically-sustainable solution will be found at scale (for example through collective licensing or an EU legislative instrument) than at the level of individual organisations, since the duplication of efforts involved in Rights Clearance for Orphan and Out-of-Print works is considerable.

7.13 Out-of-Print titles

7.13.1 Out-of-print books are printed works which are still within the period of copyright, but which are no longer actively being published by their publisher.

7.13.2 The recent Google Books Settlement (*Amended Settlement Agreement*), makes a distinction between *commercially available* and *non-commercially available* books, the latter broadly covering out-of-print works. Google is entitled to make ‘display use’ of out-of-print works without the prior permission of a rightsholder,

⁸ http://sca.jiscinvolve.org/wp/files/2009/06/sca_colltrust_orphan_works_v1-final.pdf

⁹ http://ec.europa.eu/information_society/activities/digital_libraries/doc/reports_orphan/anna_report.pdf

whereas they must make specific agreements with each rightsholder to secure permissions to make the digitised book available for display.

7.13.3 This ruling has significant implications for cultural institutions engaged in Digitisation for online public access. As we have seen previously, IPR clearances account for a significant proportion of the upfront preparation costs of the Digitisation workflow. If it can be established reliably that they are not obliged to secure permissions in respect of out-of-print works, under the precedent established by the Google settlement, then this cost can be avoided.

7.13.4 There is no way, using current evidence, to establish a reliable figure for the proportion of book titles held in European cultural institutions that are out-of-print.

7.13.5 The recent written submission¹⁰ to the Comité des Sages from the *International Federation of Reproduction Rights Organisations (IFFRO)* sets out a number of proposed basic principles for libraries wishing to digitise collections incorporating out-of-print works. These include:

- That the current status of a title as out-of-print does not necessarily indicate that publishers and authors do not intend to commercialise the work at a point in the future;
- That libraries should proceed with their Digitisation activities, but with the greatest possible transparency and in dialogue with authors and publishers;
- That Reproduction Rights Organisations (RRO) should be mandated to establish and manage licensing solutions in respect of out-of-print works.

7.13.6 The difficulty with solutions of the type presented by IFFRO is that the real-terms capital expenditure on upfront licensing of Orphan and/or out-of-print works is less palatable to cultural organisations (who do not historically have a budgetary provision sufficient to cover the proposed costs) than the potential risk of damages arising from an infringement, particularly given the relative protection afforded through risk management and takedown policies.

7.14 Digitising Rare Books in Libraries

7.14.1 So far, we have addressed the questions arising from the Digitisation of current book stock in libraries. We have not taken into account the significant quantities of rare books and similar printed matter, which account for a large proportion of their overall book holdings.

7.14.2 The acquisition, preservation and protection of rare books is an important function of National Public and University libraries, where a majority of rare book stock is held.

7.14.3 The Digitisation of rare books is expensive, due in part to the fragility and non-standard format of the material, but also because of the relative complexity of OCR scanning and particularly error-correction. For these reasons, rare-book Digitisation requires a far higher degree of human intervention, and is less susceptible to the efficiencies of mass or batch-processing.

7.14.4 Based on weighted averages, the NUMERIC report provides the following estimates for the number of rare books in 'relevant institutions' within Europe and the costs of digitising them:

¹⁰ <http://www.ifvro.org/content/ifvro-recommends-guidelines-facilitate-solutions-licensing-out-print-works>

	Number of rare books	Cost per book	Digitisation cost (€bn)
National Libraries	5,678,632	154	0.87
University Libraries	3,796,669	117	0.44
Public Libraries	12,735,200	122	1.55
Special Libraries	873,406	426	0.37

7.14.5 These figures would lead us to expect a total of 23.08m rare books in library collections, which could be digitised at a cost of 3.24bn Euro.

7.14.6 For several reasons, however, this may be an underestimation of the scale of the question. Firstly, in establishing the estimated digitisation costs, it is not clear whether participating institutions factored in *all* of the workflow costs associated with the conversion and quality-assurance of rare book material. Secondly, the NUMERIC figures relate only to a sample community of *relevant* institutions, where in fact rare book material is likely to be found in all cultural heritage institutions.

7.14.7 Further, the calculation of the likely costs of digitising rare books is far less susceptible to adjustment for duplicates, since by their very nature rare books are less frequently multiples of the same title (and even where they are, an individual volume or edition is likely to have more significance than a simple duplicate).

7.14.8 If we apply the same cost-model to the Digitisation of rare books as to current printed books, we must factor in the following elements:

- Capital costs (equipment, setup, staffing, training)
- Preparation, including unbinding (per volume cost)
- Scanning costs (per page, greyscale/colour*) x 250
- OCR conversion costs** (per page) x 250
- PDF conversion costs (per page) x 250
- Simple metadata creation (per volume)

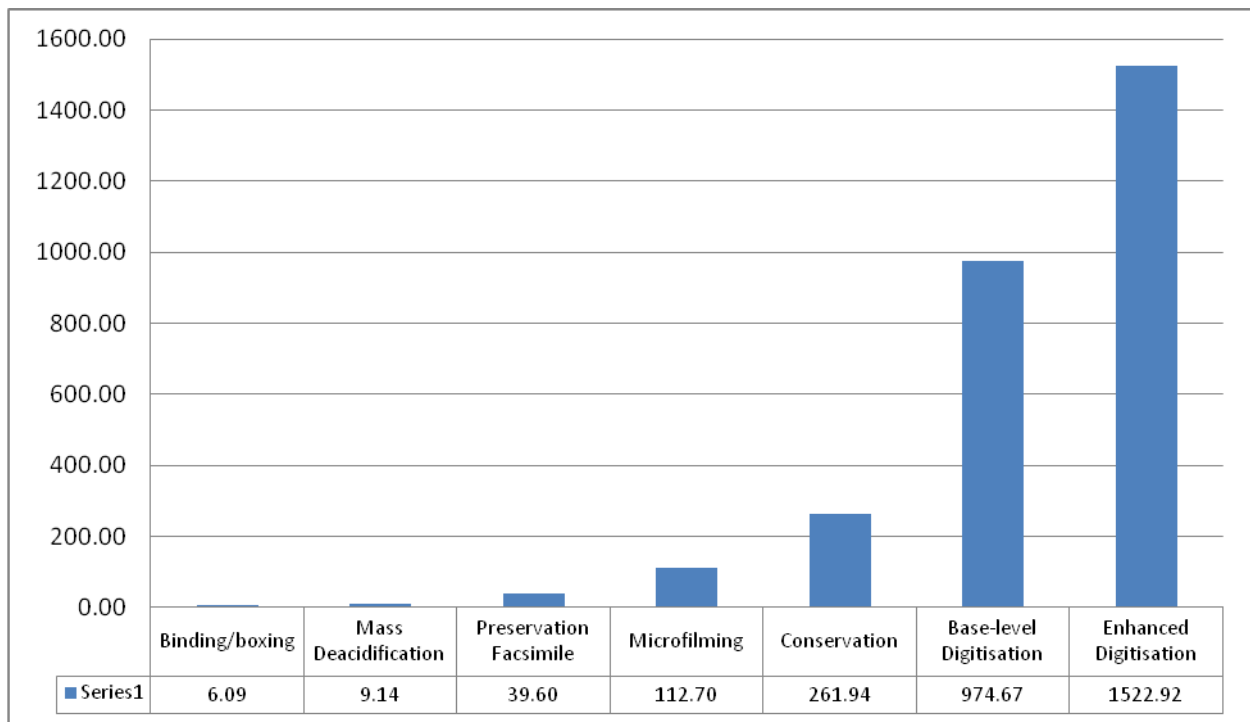
* Unlike current printed books, which in the majority of cases can be scanned to create a bitonal image without significant loss of contextual and/or cultural information, the tonality of the page is more likely to be of importance, and hence rare books would tend to be scanned in higher resolutions, with greater bit depth and in either greyscale or colour.

** As noted previously, rare books are more likely to be printed in non-standard fonts, and to be more susceptible to printing defects, which means that the OCR conversion and error-correction costs will be correspondingly much higher.

7.14.9 The Library of Congress has estimated a total of €4.26 per page for the Digitisation of a rare book, factoring in the costs of identifying and preparing materials, scanning, metadata creation, automated generation of OCR and minimally-encoded text, quality control and project management. This provides a total of €1,169 to scan a 300-page Rare Book (or €1065 to scan an average 250-page book).

7.14.10 For the costs of 'enhanced Digitisation', which creates a machine-readable text with complete SGML encoding and possible also geocoding, the costs are significantly higher - €6.10 per page, or €1,523 to scan a complete 250-page book.

7.14.11 The *Report on the Task Force on the Artefact in Library Collections*¹¹ provides the following useful chart of estimated costs for different Rare Book activities in libraries:



7.14.12 Adjusting the *quantitative* estimate of 23m Rare Books from NUMERIC is correct, then we can calculate a total range for digitising Rare Books in relevant collections (as defined under NUMERIC) as €22.42bn to €35.03bn.

7.14.13 Given an estimated 70% of Rare Books that are either too fragile to be digitised or otherwise not appropriate for Digitisation, we arrive at a compounded estimate of 6.9m books in relevant institutions at a total cost of between €6.73bn and €10.51bn (MEAN €8.62bn), depending on whether we opt for ‘base-level’ or ‘enhanced’ metadata creation and text mining.

7.14.14 It should be noted, however, that these figures are caveated on the basis of the NUMERIC selection of ‘relevant’ institutions. This selection was on the basis of several criteria, but was challenged in the recommendation report of the SIG-STATS. These extrapolated figures should be revisited in light of new evidence arising from the ENUMERATE Thematic Network.

7.15 Digitising Archival Material in Libraries

7.15.1 Libraries, and in particular National and University Libraries, hold and preserve large quantities of archival material. This material includes Government records, historical documents and other forms of archival record.

7.15.2 The following estimations are taken from the NUMERIC analysis of quantities of physical units (and resulting conversion ratios) for archival material in libraries:

¹¹ <http://www.clir.org/pubs/reports/pub103/appendix6.html>

	Metres	Pages per metre	Pages (bn)	Cost per page	Cost (€bn)
National Libraries	105082	7000	0.74	0.8	0.59
University Libraries	181280	7000	1.27	0.8	1.02
Public Libraries	42478	7000	0.30	0.8	0.24
Special Libraries	195577	7000	1.37	0.8	1.10
TOTAL					2.94

7.15.3 As before, when analysing NUMERIC data, care must be taken to allow for the fact that these figures are drawn from a sample of *relevant* institutions, and that there is some disagreement on the criteria for relevance.

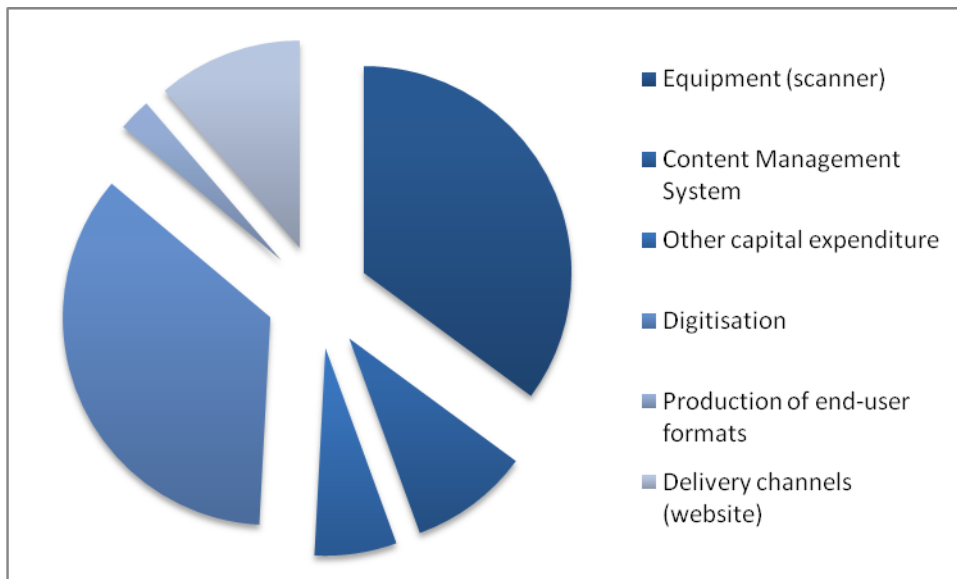
7.15.4 In particular, the per-page cost may not take into account the full costs of Digitisation workflow, which in the case of archival material may require conditional assessment and conservation, and will certainly require specific handling practices which will serve to increase the cost.

7.15.5 When compared to actual project data, certainly, the NUMERIC cost seems low. For example, the UK BOPCRIS/JISC Parliamentary Digitisation project in 2005 (see below) digitised a total of 1,260,262 pages at a cost (including equipment, staffing, scanning and metadata creation) of €1,659,053 – a cost per page of €1.32.

7.15.6 The cost ratios of the BOPCRIS/JISC Parliamentary Archives Digitisation project are revealing. The following table shows the percentage distribution of costs across this project:

<i>Item</i>	<i>Cost (€)</i>	<i>% of total costs</i>
Equipment (scanner)	582187	35
Content Management System	155298	9
other capital expenditure	106117	6
Digitisation	585594	35
Production of end-user formats	42399	3
Delivery channels (website)	187461	11

The following chart illustrates the relative distribution of costs:



7.15.7 This approximate distribution of costs (30% to equipment and capital outlay, 30% to digitisation and 30% to supporting activities and delivery) holds true across the majority of all forms of Digitisation project, except where the nature of the material is particularly complex.

7.15.8 Factoring a lower and upper estimate for the per-page cost of €0.80 and €1.32 respectively, we can calculate a lower estimate of €2.94bn and an upper estimate of €4.85bn to digitise the total Archival holdings of all relevant institutions.

7.15.9 Bearing in mind the adjustment of a 37% reduction to account for material that is not appropriate for Digitisation (cf. Section 6.2) this results in a total estimated range of €1.85 to €3.05bn (MEAN €2.45bn) to digitise archival material held in European libraries.

7.16 Digitising Newspapers in Libraries

7.16.1 Alongside books and archival material, newspapers represent one of the largest proportions of the holdings of National libraries, certainly by volume.

7.16.2 Although a significant majority of newsprint is now born-digital, the collections of libraries span newspaper publishing from the early 1700's to the present day, accounting for many millions of individual volumes and potentially billions of pages.

7.16.3 Nor is the Digitisation of newspapers an uncontested area for libraries. The May 2010 announcement of the partnership between the British Library and publishing subsidiary Brightsolid to digitise 40m+ pages from the BL Newspaper archive attracted considerable criticism from publishers and content providers concerned to preserve market value for their content offerings. It should be noted that the proposed scope of this project addresses only accounts for some 5.3% of the British Library's total holdings of 750m+ pages.

7.16.4 The per-page Digitisation costs for newspapers vary considerably depending on the fragility of the paper stock and the complexity of the format, and the other usual contingent variables attendant on the process.

7.16.5 As with all Digitisation activities, these costs are susceptible to economies of scale, as pointed out by Edwin Klijn in his 2009 article *The Quality of Quantity: Newspaper Digitisation at the Koninklijke Bibliotheek*¹². The

¹² <http://www.ifla.org/files/hq/papers/ifla75/99-klijn-en.pdf>

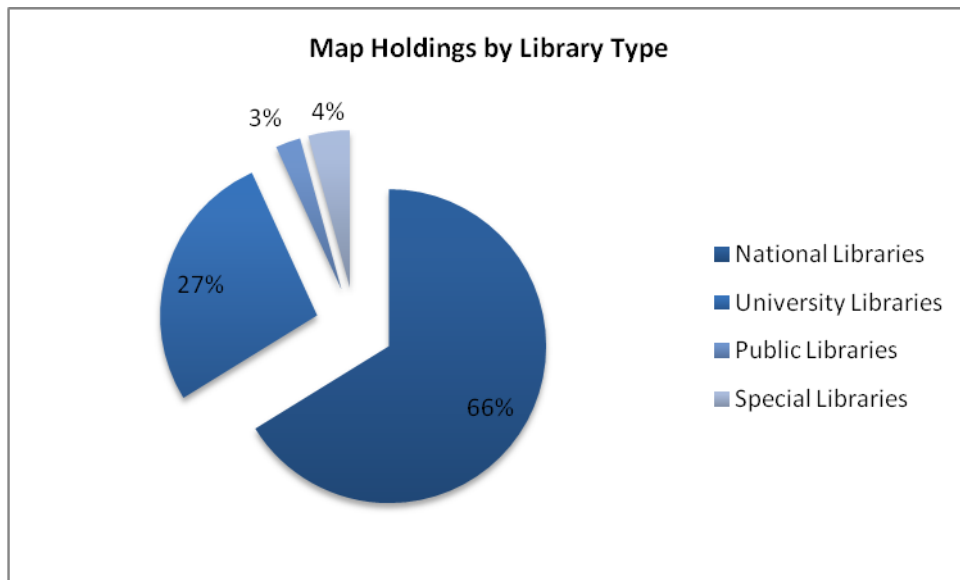
greater the throughput of material, the lower the individual per-page costs, since it becomes possible to make better use of automated equipment and to standardise metadata creation.

- 7.16.6 As with the Digitisation of books, the costs are broadly lower when the newspaper Digitisation is conducted in the context of a Public/Private Partnership than when outsourced, and both tend to be cheaper than running entirely in-house or 'boutique' Digitisation programmes.
- 7.16.7 The estimated total cost-per-page for the BL/Brightsolid project is £1 (or €1.17 on current exchange rates), a total investment of €46.8m over the 10-year period of the project. It should be noted, however, that the intended end-product of this Digitisation programme is a commercially-available content service, and that a number of the 'real' costs may be offset in this calculation against future anticipated income.
- 7.16.8 The BL/Brightsolid figure is broadly corroborated elsewhere, however. For example, in the NUMERIC findings which estimated a per-page cost of between €0.97 and €1.56, depending on the institutional context.
- 7.16.9 A more complex exercise rests in estimating both the likely total Collections of newspaper archives in Libraries, and in estimating what proportion of those Collections ought to be taken into account. This complexity is compounded by widely-reported discrepancies in the number of pages per newspaper across different editions and the quantity of material that has already undergone conversion into microforms.
- 7.16.10 This latter issue represents a very serious challenge in establishing viable data on the quantity of newspapers to be digitised. Many libraries have converted their newspaper holdings *en masse* into one or more microformats as part of earlier large-scale preservation activities, such that a majority of newspaper pages may now be in microform, rather than original format.
- 7.16.11 There is simply insufficient data to perform a valid calculation on this material, particularly since information about newspaper holdings is concatenated with information about serials, periodicals and legacy material converted into a wide variety of microforms.
- 7.16.12 Instead, then, of seeking to provide a legitimate multiplier as we have with other types of material, we would postulate a hypothetical situation in which each European Member State undertakes a large-scale Digitisation programme of its newspaper archives, along similar lines to the project of the British Library.
- 7.16.13 Given that the BL's ambition of 40m+ pages represents the largest Newspaper Digitisation undertaking to date, we would suggest establishing a median figure for our hypothetical European Digitisation Programme of 20m pages per EU Member State.
- 7.16.14 Such a programme would produce 540m pages at a total cost of between €0.52bn and €0.84bn (based on upper and lower estimates from NUMERIC and the BL – MEAN €0.68bn) including selection, preparation and base-level metadescription.
- 7.16.15 The remaining complexities in this process include regional variations in cost, and the possibility that participating nations may not have access to 20m pages of newspaper material. We have also not adjusted to account for whether these pages originate in print or microformat – since of the two, print Digitisation is the more expensive, but that relative savings on scanning costs would be balanced by increased discovery and preparatory costs.
- 7.16.16 There is no need to adjust this figure further to account for percentage exclusions, since we would anticipate that in selecting the 20m pages per EU Member State, each participant would already have excluded any material not suitable for Digitisation.

7.17 Digitising Maps in Libraries

7.17.1 The collection and preservation of cartographic material is an important function of libraries, and many libraries around Europe have been actively engaged in Map Digitisation projects over the past 10 years.

7.17.2 From the NUMERIC survey, the bulk of cartographic material is held in National Libraries (4,785,929 individual maps) and University Libraries (1,949,386). Add to this the estimated 500,000 maps in Public and Special Libraries, and we have an approximated figure of 7.23m maps in European Libraries (actual distribution shown in the chart below).



7.17.3 The costs of map Digitisation obviously vary considerably with the nature and complexity of the material, but in general it represents an expensive form of Digitisation because of the specific nature of the content.

7.17.4 Specifications for recent Map Digitisation projects tend to be at the higher end of the scale, to include:

- Hi-resolution (600-900dpi) scanning/photography
- Colour scanning/photography
- Enhanced metadata creation (including geocoding)

7.17.5 Maps also tend to incur higher handling and transportation costs, because of the wide range of non-standard formats and the need manually to open and place them, and also for this reason are less susceptible to the economies of batch-processing or mass-digitisation.

7.17.6 Because image files associated with map Digitisation tend to be very large, the ability to store the output formats while still in-project is more of an issue than with many other forms of Digitisation.

7.17.7 The costs of map Digitisation are also more susceptible than some other forms of Digitisation to significant variation on the basis of the anticipated use of the output format. It is possible to take a picture or a scan of a map that is sufficient for display, but if the intent is to create rich GIS information, Digital Elevation Models or material that can be used within CAD workflows by engineers and architects, the costs can vary from €5.92 up to €600-1000.

7.17.8 For the purposes of this enquiry, we will remain faithful to the directive to supply digital assets that are sufficient to meet the requirements of Europeana. Given that Europeana is primarily a metadata search and browse service, we anticipate that the specification would be satisfied by 'low-end' image capture of map

collections. It is important to remember that Digitisation against this relatively low specification is unlikely to satisfy the usage requirements of most end-users.

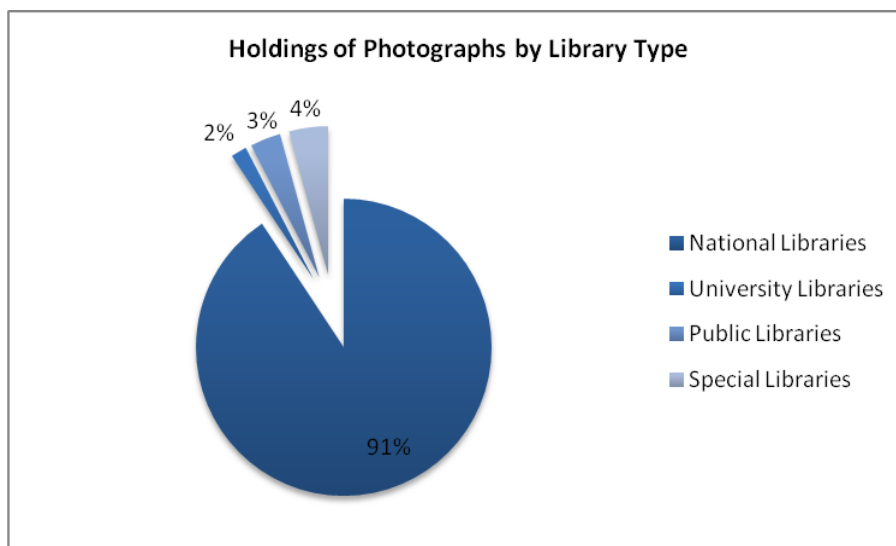
7.17.9 On this basis, however, NUMERIC provides a low threshold of €4.00 per map and a high threshold of €22.80 per map. Assuming 7.23m maps, this yields a low estimate of €28m and a high estimate of €164m.

7.17.10 Given the nature of the material, we must also assume that a significant proportion of map material is either too fragile to digitise, or is not appropriate for Digitisation. Adjusting on the assumption that this applies to 36% of material (as indicated in NUMERIC), we have a total estimated cost between €18.22m and €103.84m (MEAN €61.03m) to digitise the map collections of National and University Libraries in Europe.

7.18 Digitising Photographs in Libraries

7.18.1 Analogue photographs, slides and transparencies represent arguably the largest body of material held by Europe's libraries, archives and museums (after archival records and natural history specimen).

7.18.2 Estimates (based on the NUMERIC project, plus reported figures from National and University Libraries) put the total number of photographs held in European libraries at 34m, with the majority being held in National Libraries (see illustration overleaf).



7.18.3 Among the approximately 30m individual photographs held in National Libraries, it is estimated that some 1.2m (approximately 4%) have been digitised through previous programmes, leaving a body of 28.8m remaining.

7.18.4 It is important to remember that a significant proportion of these holdings of photographs may be ephemeral, or of relatively little cultural value, because they will have been acquired in large undifferentiated batches relating to particular collecting priorities of the institution. A significant part of any Digitisation programme will involve the selection, de-duplication and cataloguing of this undifferentiated material.

7.18.5 The UK Higher Education Data Consortium (HEDS) provides the following cost estimates for different forms of image/transparency digitisation:

Formats	150dpi 8-bit	300dpi 8-bit	600dpi 8-bit	1200dpi 8-bit	2400dpi 8-bit	300dpi 24-bit	600dpi 24-bit	2400dpi 24-bit
35mm slide					€1.74 – €2.33			€2.21 – €2.80
35mm negative					€0.93 – €1.51			€1.11 – €1.69
Other transparency sizes					€1.69 – €2.27			€2.50 – €3.09
Photographs (A3 or less)		€1.05 – €1.63	€1.30 – €1.86	€2.04 – €2.62		€2.44 – €3.03	€3.67 – €4.25	
Printed Material (A3 or less)		€0.17 – €0.29	€0.29 – €0.58	€0.52 – €1.07		€1.74 – €2.33	€2.91 – €3.49	
Photocopies (A4 or less)	€0.08 – €0.12	€0.10 – €0.15						
Glass slides and plates (A4 or less)		€11.64 – €14.55	€14.55 – €17.47					

(All prices have been converted to Euro from original GBP figures)

7.18.6 In order to establish a reasonable estimated cost for the Digitisation of photographs in libraries, we need to take into consideration the ancillary costs of:

- Selection
- De-duplication and cataloguing
- Rights clearance
- Sorting by shape/format
- Scanning
- Metadata creation
- Quality Assurance
- Project Management

7.18.7 The majority of these costs depend on staff time, and hence can be extremely difficult to calculate on a normalised basis. The overall cost is a function of the staff costs and the work rate of the individual staff, noting that work rate tends to improve over time in larger projects (as workflows are refined and staff gain in competence).

7.18.8 Normalised estimates from a range of medium to large-scale projects indicate a range of:

- €4-6 per image for simple, straightforward images
- €9-12 per image for non-standard, damaged or complex images
- €12-15 per image for oversized or very complex images

7.18.9 If we were to proceed with the Digitisation of all 28.8m photographs estimated to be held in European libraries, the costs would be very considerable (see table below):

<i>Material type</i>	<i>Estimate</i>	<i>Per-image cost (€)</i>	<i>Total estimated cost(€m)</i>
Simple	Low	4	115.2
	High	6	172.8
Complex	Low	9	259.2
	High	12	345.6
Oversized	Low	12	345.6
	High	15	432

7.18.10 Adjusting on the basis that libraries have indicated that approximately 70% of their collections are either not suitable or not eligible for Digitisation, we have the following revised estimates, based on 8.64m photographs (30% of the total estimated un-digitised holdings):

<i>Material type</i>	<i>Estimate</i>	<i>Per-image cost (€)</i>	<i>Total estimated cost (€m)</i>
Simple	Low	4	34.56
	High	6	51.84
Complex	Low	9	77.76
	High	12	103.68
Oversized	Low	12	103.68
	High	15	129.6

7.18.11 If we calculate a rough estimate of 60% of holdings as ‘simple’, 20% as ‘complex’ and 10% as ‘oversized’, we can estimate the total cost range between €14m and €19.44m to digitise 8.64m photographs across European libraries.

7.18.12 Of all of our estimates, this one is perhaps prone to the greatest margin of error – it is only by investigating the collections to identify the nature of the material and the extent of the duplication that we would be able to arrive at a more accurate estimate. We therefore recommend that considerable additional research is done in this particular area to improve our overall knowledge.

7.19 Digitising AV material in Libraries

7.19.1 Given the specific nature of Audio Visual material, and the particular cost implications, we have investigated the cost implications of digitising Europe’s AV collections overall in a later section.

7.20 The Cost of Digitising Libraries

7.20.1 We are now in a position to evaluate the overall estimated costs of digitising different types of material held in library collections. The schedule of costs below shows the main components of this calculation:

<i>Material</i>	<i>Type of Digitisation</i>	<i>Lower estimate</i>	<i>Higher estimate</i>
Books	In-house	7.32	16.2
	Outsourced	4.32	9.83
	Public Private Partnership	3.25	6.32

Rare books (incl. manuscripts & incunabula)		6.73	10.51
Archives		1.85	3.05
Newspapers		0.52	0.84
Maps		0.018	0.103
Photographs		0.014	0.19

7.20.2 Mapped against our 3 proposed ‘models’ for Digitisation (in-house, project, based Digitisation versus outsourced Digitisation or work completed under a Public Private Partnership), we can therefore calculate the following estimates:

	Lower estimate (€bn)	Higher estimate (€bn)	MEAN estimate (€bn)
Distributed, predominantly in-house Digitisation	16.45	30.89	23.67
Distributed, predominantly outsourced Digitisation	13.45	24.52	18.95
Public-Private Partnership Digitisation	12.38	21.01	

7.20.3 As noted before, the cost differential between the models masks the greater complexity of the way in which the real costs of Digitisation are allocated across different project activities, and they are not directly comparable on this basis.

7.20.4 It should be noted that the cost of digitising Audiovisual material has been explicitly excluded from this calculation. These costs will be factored into the final calculation of overall costs.

7.20.5 There is no accurate way of calculating Digitisation costs outside of the context of an actual project. These figures are based on a series of projections, estimates and extrapolations from real costs in current projects. They should be read as indicative only, and all are subject to a margin of error.

8. The Cost of Digitising Museums

8.1 Scope

- 8.1.1 Digitisation in Museums is different from Mass-Digitisation in Libraries. As noted previously, it is less a question of converting cultural material from one format (such as a printed book) to another, and more of creating a digital *surrogate* of the original physical artefact for the purposes of promotion, access and conservation.
- 8.1.2 The museums sector is smaller than the library sector in Europe and although in some areas the scope of collections is comparable (such as natural science), in most areas museums are dealing with a smaller quantity of material.
- 8.1.3 Because museum digitisation is concerned partly with the communication of narrative and rich cultural information, the attendant costs are much higher than for the Digitisation of books. Also, because there is a smaller quantity of material, museums are less able to establish and refine the type of large-scale workflows which lead to significant efficiency savings in libraries.
- 8.1.4 Allied to this is the greater diversity of object types in museum collections. A museum's collections may span artefacts from the microscopic to the gigantic, covering a range of man-made and natural artefacts. While books can be unbound and fed through sheet-feeding scanners, objects in museum collections may require specific equipment, or may need to be photographed from multiple angles to create a suitable image.
- 8.1.5 Not only are museum objects generally more complex in terms of scanning and reproduction, they are less susceptible to batch-processing for metadata-creation (with the exception of well-documented collections of slide transparencies). Hence while book scanning can produce machine-readable text via OCR, all of the metadata associated with an individual artefact must be created through human intervention.
- 8.1.6 It is important, when considering museums, to differentiate between larger-scale National Museums (akin to National Libraries and Archives) and the long-tail of smaller institutions, many of which are either commercially independent or supported at a local or regional Government level.

8.2 Output Formats for Museum Digitisation

- 8.2.1 Digitised images of objects in museum collections can be used in a number of ways:
- Hi-resolution scans to support conservation and conditional assessment
 - Low-resolution images for delivery through online collections databases
 - Low-resolution images to illustrate records in Collections Management Systems
 - Inclusion in catalogues and posters for the purposes of marketing and promotion
 - Images of various resolutions for image licensing, retail and art-on-demand services
- 8.2.2 Not only this, but museums are increasingly investigating the possibility of outputting more complex formats, such as high-resolution 3-dimensional renders to be inserted into special effects workflows for television production, or scans to provide templates for 3D printing and fabrication.
- 8.2.3 Because of the potential promotional and commercial use of images of museum objects, museums tend to expend more effort to create digital assets that are higher-value, more attractive and better-lit (leading one National museum to differentiate between 'Digitisation' and 'beautiful Digitisation'), and to spend more time on the preparation of individual objects for photography.

8.2.4 Many museums use the opportunity of photographing an object to create more than one image type, to satisfy different use requirements. As an example, the V&A museum in London creates the following 3 'types' of photograph from a single session¹³:

- i) A *record shot*, usually taken by a non-specialist photographer, to be used for general purposes such as exhibition planning;
- ii) A *creative shot* in which the object is photographed in an attractive setting, for uses in posters and promotion; and
- iii) A *descriptive image*, which is taken at the highest available resolution and used for long-term curation, preservation management and identification.

8.3 Scale of Museum Digitisation

8.3.1 Due to these considerations, the average workrate, or output rate of museum Digitisation projects tends to be slower than in a large-scale, batch-processing library Digitisation project.

8.3.2 Also, due to the different nature of the material, museums have not attracted interest from commercial Digitisation partners in the same way as libraries. There is no equivalent to the Google Book Search or Proquest Digitisation projects in the museum community, with the result that there has been less technical innovation in the field and a lower overall investment in museum Digitisation.

8.3.3 The NUMERIC project indicated that European museums estimate that a higher percentage of their Collections is valid for Digitisation (only 3% is excluded, as opposed to 36% in archives and 69% in libraries), and also that a significantly higher proportion of their Collections (25%) have already been digitised.

8.3.4 Even allowing for the potential of a slight bias on the basis of the profile of museum participants in NUMERIC (generally, larger, better-equipped National museums), this figure is proportionately higher when compared to libraries and archives, most likely due to the significantly lower overall quantity of material held in museums.

8.3.5 It is important to note at this point the significantly smaller quantity of available research data on the subject of specific Digitisation costs in museums. This seems likely to be due to the quantity of Digitisation activity that is undertaken in the context of other activities (such as conservation), but very few museums in Europe publish explicit data about their Digitisation expenditure as distinct from other collections-related activities.

8.4 Museum Collections in Europe

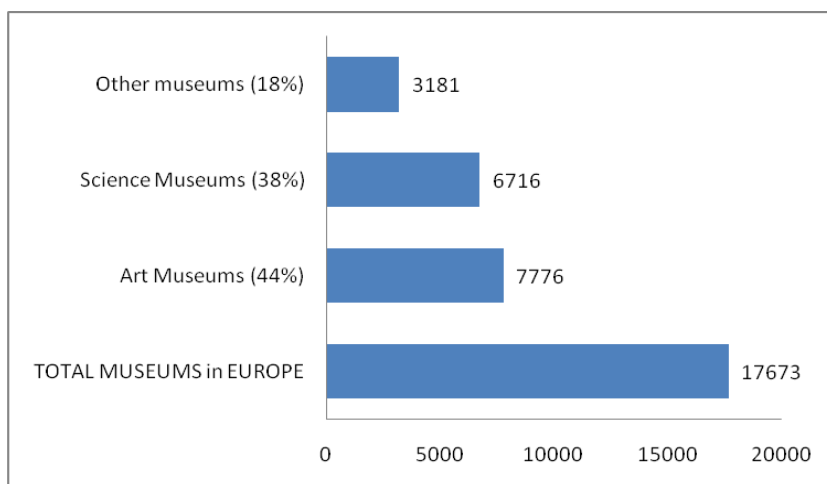
8.4.1 There are approximately 17,673 museums in Europe, based on figures submitted between 2003 and 2008¹⁴. There are, on average 0.000080 museums per capita of population in the European Member States, averaging out at one museum for every 33,143 European citizens.

8.4.2 Of the c. 17,673 museums in the EU, we estimate the following divisions between art historical/archaeological museums, natural science museums, and other types of museum¹⁵:

¹³ Source: *Digitisation Programmes in the V&A* (James Stevenson)

¹⁴ Extrapolated from European Commission published figures & reports provided by the EGMUS network (see appendix)

¹⁵ Based on figures provided by the NUMERIC final report (see appendix)



8.4.3 The distinctions between museum types are statistically significant because of the broad differences in the nature of their Collections. Similarly, the difference between the holdings of the large number of small museums and the relatively small number of very large museums is also a significant statistical factor.

8.4.4 Where the predominant material types in libraries were books, archival material, photographs and other printed matter, the predominant types in museums are:

- Man-made artefacts
- Naturally-occurring material
- 2D works of art
- 3D works of art
- Photographs

8.4.5 We should note, from the outset, that these classifications cover an incredibly diverse range of material. ‘Man-made’ artefacts, for example, are likely to include everything from complex machinery to wooden furniture, from fragile ancient pottery to contemporary plastics. There is clearly no statistically-legitimate way to normalise the costs of Digitisation across this range of material. To address this, we will average out per-object costs on the basis of existing projects. The resulting figures should be taken as no more than very broadly indicative, and we anticipate that future projects will differentiate more specifically between objects of different material types.

8.5 Cost Models for Museum Digitisation

8.5.1 In essence, the components of the basic cost model for museum Digitisation are the same as those for libraries and archives, comprising:

- Selection/preparation
- Conservation
- Rights clearance
- Location & movement control
- Scanning/photography
- Metadata creation
- Quality Assurance
- Project Management

8.5.2 The models for Digitisation projects tend to be similar to those for libraries, but at a smaller scale, and without the option of Public Private Partnership Digitisation at any scale other than the larger National museums. They include:

- Small-scale ‘boutique’ Digitisation projects undertaken in-house and based on particular thematic collections or groupings of object/material type.
- Larger-scale Digitisation projects undertaken either on an outsourced-based, or in-sourced (for example, using mobile Digitisation labs or onsite scanning facilities operated by 3rd-party contractors)
- Ongoing Digitisation (photography and scanning of slides and transparencies) as part of the normal workflow of Collections documentation, conservation and cataloguing.

8.5.3 Many museums maintain facilities for the creation of digital surrogates of objects as they move through the normal workflow of Collections Management. In his 2004 study of *Reproduction Charging Models* on behalf of the US-based Mellon Foundation¹⁶, Simon Tanner of Kings Digital Consultancy Services (KDCCS) highlights the broad divisions of investment in Digitisation infrastructure shown overleaf.

8.5.4 It is important here to distinguish between several different ‘strands’ of Digitisation activity in museums. In particular, the ongoing process of conversion of slide transparencies (themselves the result of earlier programmes to record collections photographically) is different from the Digitisation of larger objects, which in turn is different from the production of high-value images for retail or promotional purposes.

<i>Type of Digitisation infrastructure</i>	<i>% of sample museums</i>
In-house Digitisation facilities (owned by museum)	66%
Outsourced Digitisation infrastructure (hosted in museum)	9%
Mixture of both	8%
No Digitisation infrastructure	17%

Table: Proportion of sample of US museums by Digitisation infrastructure (Tanner, 2004)

8.5.5 These figures raise an interesting possibility. It may be that long-term strategic investment in building core capacity in museums to include Digitisation as part of ongoing core activity delivers a greater return on investment than supporting Museum Digitisation through challenge or project funding, since in the long-run, the unit cost of Digitisation will be reduced. There is insufficient evidence to justify this hypothesis, but it may prove an interesting avenue of enquiry for future research.

8.6 Digitising Man-made Objects in Museums

8.6.1 Since man-made artefacts make up a significant proportion of the overall holdings of museum collections, we will begin our investigation with these.

¹⁶ http://www.kdcs.kcl.ac.uk/fileadmin/documents/USMuseum_SimonTanner.pdf

8.6.2 There is very little evidence concerning the total quantities of objects of any type, including man-made objects, in museum collections in Europe. Relatively little mapping has been completed, and where it has, it tends to be regionally-specific and not applicable across the broader Eurozone.

8.6.3 The NUMERIC project provides the following estimates of the quantities of man-made material in museum collections:

<i>Museum type:</i>	<i>Number of man-made objects (median)</i>	<i>Number of museums</i>	<i>Estimated number of objects</i>
Art/Archaeology/History Museums	17476	7776	135893376
Science/Technology Museums	14200	6716	95367200
Other types of Museum	10501	3181	33403681
TOTAL			265m

8.6.4 Across all museum types, therefore, NUMERIC indicates an approximate figure of 265m for the total number of man-made objects in their collections. This is equivalent to a mean average of 14,976 man-made objects per museum.

8.6.5 This projected figure excludes all natural materials, archives, audiovisual material, paintings, sculpture and photography.

8.6.6 As noted above, the costs of digitising man-made artefacts varies considerably according to the material in question. NUMERIC provides an estimated scanning cost per unit of between €3.92 and €6.68, but this must be set in context of the related costs of preparation, processing and description.

8.6.7 The calculus of the distribution of funds across related project activities differs from the split in libraries into fairly equal 3rds, illustrated by the exemplar project below (figures from a small museum in-house Digitisation programme)¹⁷:

<i>Cost centre</i>	<i>Actual cost (€)</i>	<i>% of total costs</i>
Staff costs	80500	44%
Overheads	14559	8%
Equipment & software	86961	47%
TOTAL cost	182021	
Number of objects digitised	1500	
Cost per object	€121	

¹⁷ Source: JISC Digitisation reports

- 8.6.8 Within this type of cost model, the actual cost of the specific act of scanning or photographing the artefact is marginal when compared to the attendant costs of management, preservation, rights clearance and other supporting activities.
- 8.6.9 Based on figures from NUMERIC and related project analysis, we can estimate an averaged-out cost-per-object of €25 to €136 (MEAN €80.50) to digitise fully a man-made artefact (noting that the spread of potential costs reflects the broad spectrum of material types in these collections).
- 8.6.10 Given that museums identify 28% of their collections as either having already been digitised or not being suitable for Digitisation, we have an estimated 191m candidate objects suitable for Digitisation.
- 8.6.11 Our total range of costs, therefore, to digitise these artefacts is between €4.76bn and €25.92bn (MEAN €15.34bn).

8.7 Digitising Natural Materials in Museums

8.7.1 Natural materials in museum collections include a wide range of material including:

- Flora (including type specimen)
- Fauna (including type specimen)
- Geological material (including chemical and mineral material)

8.7.2 Based on estimates, we can extrapolate the following median holdings of natural materials in museums:

<i>Museum Type</i>	<i>Number of natural objects (median)</i>	<i>Number of museums</i>	<i>Estimated number of objects</i>
Art/Archaeology/History Museums	93	7776	723168
Science/Technology Museums	37359	6716	250,903,044
Other types of Museum	17301	3181	55,034,481
TOTAL			307m

- 8.7.3 Across all museum types, there are an estimated 307m natural science objects, giving an average of 17,351 objects per museum.
- 8.7.4 Adjusted to account for the 28% of material either already digitised or not suitable for Digitisation, we estimate an approximate total of 221m naturally-occurring objects eligible for Digitisation in European museums.
- 8.7.5 Because natural science materials occupy a similarly broad spectrum to man-made materials, the range of potential unit costs is correspondingly broad – ranging from €26 to €121 per unit (and adjusted to account for anomalously large, complex, or hazardous objects).
- 8.7.6 The range of total costs is therefore very considerable, ranging between €5.74bn and €26.72bn (MEAN €16.23bn) to digitise the total holdings of 221m objects (a median unit cost of €73 per individual object).

8.8 Digitising Works of Art in Museums

8.8.1 Although works of art (both 2D and 3D) are relatively low in numbers, they represent a significant part of the holdings of European museums and present a particular challenge in terms of Digitisation.

8.8.2 As with other forms of man-made object, the costs of digitising Works of Art mainly comprise staffing and overhead costs of preparation, conservation and metadata creation. The actual Digitisation cost *per se* is a relatively minor part of the overall cost.

8.8.3 These costs are compounded by the relative complexity of moving works of art so that they can be digitised. Movement of works frequently requires expert technical staff, and will have a considerable impact on the work or throughput rate of the Digitisation facility (which is why many museums elect to digitise objects *in situ* rather than move them).

8.8.4 For the purposes of this investigation, we will consider the following:

- Paintings
- Sculptures
- Prints
- Drawings
- Engravings

8.8.5 NUMERIC provides us with the following estimates of total holdings of each of these material types in different types of museum:

	<i>Art/ Archaeology (m)</i>	<i>Science/ Technology (m)</i>	<i>Other (m)</i>	<i>TOTAL (million objects)</i>	<i>Equivalent to....</i>	
Paintings	5.83	0.67	1.11	7.62	431	paintings per museum
Sculptures	6.22	0.94	2.54	9.71	549	sculptures per museum
Posters/postcards	6.61	9.16	2.86	18.63	1054	posters/postcards per museum
Drawings	7.37	20.15	1.44	28.96	1639	drawings per museum
Engravings/prints	7.08	1.52	1.91	10.51	595	engravings/prints per museum

8.8.6 Based on these figures, we can estimate a total of 75.43m works of art in European Museums, equivalent to 4,268 works for every museum in Europe.

8.8.7 The unit costs of Digitisation for works of art are very similar to those for digitising other forms of man-made objects, because the processes themselves are essentially the same (not including any significantly anomalous costs such as the need to conserve a particular work). Hence, we can presume a similar scale of €25 to €136 (MEAN €80.50) for the Digitisation of works of art.

8.8.8 An important consideration, however, is that a significantly higher proportion of eligible works of art will already have been digitised, since these are frequently either ‘masterpieces’ or ‘star’ items in collections. Many Art Museums, for example, already hold complete or near-complete digital records of their 2D works.

8.8.9 There is no evidenced mechanism for estimating the proportion of works of art in public ownership that have been digitised. And hence it is very difficult to establish what proportion to exclude from this calculation. If we assume a total exclusion of 28% (as with other types of museum collection), then our total number of candidate works is 54.30m, giving a projected cost of €1.36bn to €7.39bn.

8.9 Digitising Photographs in Museums

8.9.1 As with libraries and archives, museums hold very significant quantities of photographic material. Similarly, too, it is extremely difficult to formulate reliable estimates due to the considerable amount of material that is either uncatalogued or catalogued at ‘box’ rather than ‘item’ level.

8.9.2 With this caveat, NUMERIC provides the following estimates for museum holdings of photographic material:

<i>Museum Type</i>	<i>Number of photographs (median)</i>	<i>Number of museums</i>	<i>Estimated number of objects</i>
Art/Archaeology/History Museums	10,000	7776	77.76m
Science/Technology Museums	53,600	6716	359.98m
Other types of Museum	15000	3181	47.72m
TOTAL			485.45m
72% considered eligible for Digitisation			350m

8.9.3 On first consideration, a total of 485.45m photographs in museum collections in the EU Member States seems to be an extraordinarily high estimated, but in practice, this equates to approximately 27,500 photographs per museum. Given the extensive photographic archives held by some museums, this chimes with normal expectations.

8.9.4 Employing the same cost ranges as for the Digitisation of photographs in libraries yields the following calculation:

<i>Material type</i>	<i>Estimate</i>	<i>Per-image cost (€)</i>	<i>Total estimated cost(€m)</i>
Simple (60%)	Low	4	0.84
	High	6	1.26
Complex (20%)	Low	9	0.63
	High	12	0.84
Oversized (10%)	Low	12	0.42
	High	15	0.52

8.9.5 In aggregate, this model gives a total estimated cost for the Digitisation of the eligible photographs in museum collections at between €1.89b and €3.67bn.

8.9.6 Great care should be taken, however, in considering these figures. They do not, for example, exclude multiples and are based on very broad estimates about the proportion of a collection that will fall into each category.

8.10 The Cost of Digitising Museums

8.10.1 We are now in a position to evaluate the overall estimated costs of digitising different types of material held in museum collections. The schedule of costs below shows the main components of this calculation:

<i>Material</i>	<i>Lower estimate (€bn)</i>	<i>Higher estimate (€bn)</i>
Man-made Objects	4.76	25.92
Natural material	5.75	26.72
Works of Art	1.36	7.39
Photographs	1.89	3.67

8.10.2 Using these aggregated figures, we can now calculate an estimated cost of digitising the eligible material in museums:

	<i>Lower estimate (€bn)</i>	<i>Higher estimate (€bn)</i>	<i>Mean (€bn)</i>
Digitising museum collections	13.75	63.7	38.73

8.10.3 It should be noted that the cost of digitising Audiovisual material has been explicitly excluded from this calculation. These costs will be factored into the final calculation of overall costs.

8.10.4 As previously noted, the max/min range for museum Digitisation is very high (almost €50bn). This is unfortunate in the broader aim of this investigation, but perhaps not surprising in light of the breadth of material under consideration, and the relatively poor available data with which to refine the model.

8.10.5 It appears from this study that museums do not regularly keep or publish detailed information about their Digitisation activities, and that, barring externally-funded projects, these costs tend to be allocated across existing internal functions. We would recommend working with museum Funding Agencies and statistical efforts such as EGMUS and ENUMERATE to refine the cost model for this area, and to ensure that data about Digitisation and related activities are being captured alongside national survey efforts.

9. The Cost of Digitising Archives

9.1 Archival Collections in Europe

9.1.1 For the purposes of this report, an 'archive' is defined as,

*Organisation or part of an organisation responsible for selection, acquisition, preservation and availability of one or more archives*¹⁸.

9.1.2 There is no definitive source of statistically-valid information about the quantity or distribution of archives within the EU, nor of the total holdings or archival material in Cultural Heritage institutions.

9.1.3 Because of the very great diversity of archival holdings, the proportion of material that is un-accessioned at any given moment and the ongoing process of selection and de-selection, it is unlikely that it will ever be possible to provide a truly accurate figure for the cost of digitising this material.

9.1.4 The figures in this section should therefore be taken as *broadly indicative*, based on our best estimates and the figures currently available.

9.2 How Many Archives are there?

9.2.1 For the purposes of this study, we have defined the scope of our interest as being:

- National Archives
- Local and Provincial Archives
- University Archives
- Archives of Foreign Ministries (Ministries of Foreign Affairs)¹⁹

9.2.2 Given the relative lack of data (and also the fact that reported figures for National Archives also cover their custodianship of Local and Provincial Archives), we have focussed on the reported holdings of the National institutions.

9.2.3 We have specifically excluded from the scope of our calculations:

- Business Archives
- Community Archives
- Church Archives
- Private Archives
- Other 'Special' archives

9.2.4 These have been excluded on the basis that (a) insufficient data is available to quantify their number or the scope of their collections and (b) they are likely in any case to adopt different approaches to Digitisation, based on funding support from other sources.

9.2.5 For the sake of ease of interpretation, we have factored into these calculations likely estimates of the quantity of different types of archival material held in museums and libraries.

¹⁸ Source: ISO 5127 'Info and Documentation: Vocabulary' - http://www.iso.org/iso/catalogue_detail.htm?csnumber=33636

¹⁹ Published by the Council of Europe – <http://consilium.europa.eu>

9.2.6 Finally, due to the significant difference in distribution and costs, we have excluded from the scope of this section Audio Visual Archives, which will be addressed in a separate section of this report.

9.3 Types of Material held in Archives

9.3.1 Archives maintain Collections of a very wide range of different material types, including²⁰:

- Laws
- Decrees
- Ordinances
- Rules
- Regulations
- Instructions
- Circular letters
- Orders
- Books
- Reports
- Protocols
- Accounts
- Plans
- Budgets
- Agreements
- Contracts
- Official and private correspondence
- Manuscripts of scientific and artistic works
- Maps
- Memoirs
- Newspapers
- Diaries
- Periodicals
- Posters
- Placards
- Scientific-technical documentation
- Photographs
- Sound-recordings
- Wills
- Movies
- Electronic and other documents in the original or copies.

9.3.2 A simplified list would include:

- Legal records
- Personal records
- Company/institutional records
- Photographs
- Audiovisual Material (film and sound)
- Electronic records

²⁰ Source: National Archive of Bulgaria - <http://www.archives.government.bg/index.php?lang=en&page=12>

9.3.3 The published data about the holdings or archives in the EU tends only to differentiate along very broad lines – hence holdings of legal records and personal correspondence tend to be concatenated into ‘fonds’, usually expressed in terms of linear meters of shelving.

9.3.4 For this reason, instead of focussing on detailed granular costings for different material types, we have focussed on the following broad sub-divisions:

- Archival records (in linear metres of shelving)
- Photographs
- Microforms (used as finding aids)
- AV material (dealt with in a separate section)

9.4 Measuring Shelving

9.4.1 Most published data about archival holdings is given in *linear metres of shelving*. This is an inexact measurement arrived at through custom and practice in archives, and it reflects the difficulty of quantifying archival holdings when these can include many hundreds of pages per metre.

9.4.2 The linear metre of shelving has been adopted as the unit of measurement for this study, and in order to establish reasonable approximations on cost, we have adopted the following multipliers, taken from the original NUMERIC report:

<i>Unit</i>	<i>Quantity</i>
Pages per linear metre (historical archives)	5991
Pages per linear metre (Government archives)	7000

9.4.3 Where possible, we will use the specific multiplier for each record type. Where record types have not been indicated (in most cases), we will presume an average of 6000 pages per linear metre.

9.5 Defining Digitisation in Archives

9.5.1 In order to achieve even a reasonable estimation of the likely costs of digitising non-Audiovisual Archival material in EU archives, it is important to consider a number of factors:

- The nature of ‘Digitisation’ and description in archives
- The extent of previous programmes of description
- The particular impact of born-digital material on archives

9.5.2 As with Digitisation in libraries and museums, the word tends to refer to a number of processes by which material is converted or photographed in a digital format. However, in archives, the term ‘Digitisation’ is used to refer to a range of quite distinct activities, including:

- Selection, preparation and preservation of material
- Prioritisation of material (based on historical or ‘informational’ value)
- Creation of a digital surrogate (photograph or scan)
- OCR interpretation of text
- Creation of catalogue, item, or collections-level metadata
- Provision of ‘meta’ metadata (such as a METS record)

- 9.5.3 In others words, Digitisation in an archival context is often a hybrid concept combining features of digital *conversion* of library holdings (periodicals, books, newspapers), digital *surrogacy* of historical records and long-term Digital Preservation of materials.
- 9.5.4 Archives and libraries discovered the potential of electronic finding aids such as the *Online Public Access Catalogue* (OPAC) some time before museums, and hence the progression of *description* and to some extent Digitisation is more advanced in archives than elsewhere.
- 9.5.5 For this reason, a significant quantity of archival material is identified as either having already been digitised, or as not requiring Digitisation (due to the unsuitability of the medium or the availability of a comprehensive electronic record. Hence, NUMERIC provides the following indications of the quantities of archival material already digitised as part of previous programmes (or already available as born-digital material):

<i>Archive type</i>	<i>Proportion already digitised (%)</i>
Government records	50
Historic records	25
Other types of archival material	54

- 9.5.6 As with libraries, the exponential increase in the proportion of new material that is born-digital is having a profound impact on future Digitisation activity in the archives sector. The majority of new material is born-digital, with the result that Digitisation is primarily a legacy concern in archives, focussing on the conversion, management and description of paper materials acquired during the previous century.
- 9.5.7 This new material may already be available in a digital form, but it nevertheless requires a significant investment in curation, management, description and long-term preservation. The archiving of born-digital material, including web archiving, is outside the remit of this investigation. It nevertheless represents a significant long-term commitment on the part of archives throughout Europe.

9.6 National Archive Collections

- 9.6.1 The table overleaf shows estimated holdings for each type of archive in each of the 27 EU Member States.

Countries	National Archives (linear metres) ²¹	BnPages (based on 6000 pages/metre)
Austria	177700	1.07
Belgium	200000	1.20
Bulgaria	3000	0.02
Cyprus	4453	0.03
Czech Republic	95000	0.57
Denmark	140000	0.84
Estonia	73000	0.44
Finland	180000	1.08
France	364000	2.18

²¹ Based on reported figures from each National Archive (dates 2005-10)

Germany	300000	1.80
Greece	8000	0.05
Hungary	71000	0.43
Ireland	Data unavailable	0.00
Italy	1600000	9.60
Latvia	13500	0.08
Lithuania	67500	0.41
Luxembourg	32000	0.19
Malta	10000	0.06
Netherlands	93000	0.56
Poland	220000	1.32
Portugal	6000	0.04
Romania	Data unavailable	0.00
Slovak Republic	157000	0.94
Slovenia	1300	0.01
Spain	220000	1.32
Sweden	300000	1.80
United Kingdom	160000	0.96

- 9.6.2 These figures provide a total estimated number of pages of archival material in Archives around the EU of 26.98bn. Of this, an estimated 36% (NUMERIC) is not suitable for Digitisation, giving a total number of candidate records of circa. 17.27bn pages (adjusted to allow for unreported holdings in the National Archives of Ireland and Romania).
- 9.6.3 Bearing in mind that most National Archives around the EU have active Digitisation programmes, the average proportion of archival holdings already digitised ranges between 25% and 54%. This means that the estimated number of pages in National Archives that are suitable for Digitisation and that have not yet been digitised ranges between 7.94bn and 12.95bn (MEAN 10.45bn).
- 9.6.4 It is important to note that these figures are based on an average estimate of 6000 pages per linear metre of shelving, which may be an overestimation of the real figure (depending on the type of material stored).

9.7 The Per-page Cost of Digitising Archives

- 9.7.1 As with libraries and museums, the per-page Digitisation costs in archives can vary significantly depending on the organisational infrastructure, the skill of the people involved, and the scale of the operation.
- 9.7.2 At the lowest end, archives have reported a per-page *scanning* cost (noting that this does not include allocations for overhead, equipment, metadata enhancement or longer-term management) of between €0.36 and €0.46.
- 9.7.3 Taking into account the full economic costs, factoring overhead and equipment/processing costs, the per-page Digitisation costs for archival material are similar to those for photography, at around €4.00 per page, depending on the organisational context and the fragility/complexity of the material.

9.8 Total Cost of Digitising National Archives

- 9.8.1 The provision of archives is not standard across Europe, and many of the National or State archives are also umbrella organisations for larger communities of federal or provincial archives.

9.8.2 On this basis, we can prepare the following schedule of costs for digitising the archival holdings of National Archives:

<i>Estimated number of pages</i>	<i>Total estimated cost</i>
7.94bn	€31.77bn
12.95bn	€51.80bn

9.8.3 This yields a MEAN figure of €41.79bn to digitise archival material in the National Archives. To this, we must add the estimated totals for the holdings of photographic and microform material in archives.

9.8.4 NUMERIC estimates a total of 692,908 units of microfilm per archive institution. Across the 25 National archives included in this study, this would account for approximately 17m units. At an estimated *scanning* cost per unit of €0.23, this would amount to a total of €3.91m. Based on a 1:1:1 ration between scanning, overhead and enhancement (see earlier), this would approximate a total of €11.73m attributable to microform scanning.

9.8.5 Photographic collections account for a significant proportion of the holdings of National archives. Again, NUMERIC estimates an average of 331,788 photographs per archive institution. Across the National archives, this would account for approximately 8.29m photographs.

9.8.6 Factoring in the same estimated cost ranges for the Digitisation of photographs in archives as in libraries provides the following estimated range:

<i>Material type</i>	<i>Estimate</i>	<i>Per-image cost (€)</i>	<i>Total estimated cost(€m)</i>
Simple	Low	4	48.29
	High	6	68.29
Complex	Low	9	98.29
	High	12	128.29
Oversized	Low	12	128.29
	High	15	158.29

9.8.7 Adjusting these figures to account for the 36% of archival holdings judged not to merit, or to be capable of Digitisation yields the following:

<i>Material type</i>	<i>Estimate</i>	<i>Per-image cost (€)</i>	<i>Total estimated cost(€m)</i>
Simple	Low	4	30.9056
	High	6	43.7056
Complex	Low	9	62.9056
	High	12	82.1056
Oversized	Low	12	82.1056
	High	15	101.3056

9.8.8 Hence the estimated MEAN figure for the digitisation of photographic material in the National archives would be €64.51m.

9.8.9 These figures provide a (very rough) estimated cost of digitising the photographic, archival and microform holdings of the 25 National libraries in the EU (excluding Ireland and Romania, but including their branches and service points) at:

<i>Material</i>	<i>Median cost</i>
Archival records	€41.79bn
Photographic collections in archives	€64.51m
Microform collections in archives	€11.73m
TOTAL	€41.87bn

10. The Cost of Digitising AV Material

10.1 About Audiovisual Collections

- 10.1.1 Audiovisual material represents a significant and growing part of the legacy of the past 100 years of social and technological development.
- 10.1.2 'Audiovisual' collections cover a wide variety of material in a number of formats including:
- Sound recordings
 - Recorded music
 - Scientific and medical material
 - Moving images on film
 - Video material
 - Broadcast TV footage
 - Animation
 - Digital movie files
- 10.1.3 Audiovisual material occurs in many different types of institution, including museums, archives, libraries, broadcasters and specialist repositories of moving images and sound. Although much of the more recent audiovisual output is born-digital, the legacy of the past 100 years includes millions of hours of material which requires conversion into digital formats for the purposes of long-term preservation and re-use.
- 10.1.4 The challenge of digitising the rich collections of audiovisual material throughout Europe is immense. The material itself is stored on multiple formats, many of which are fragile and prone to rapid decomposition. The process of Digitisation frequently requires the maintenance and use of outdated equipment. The material itself requires rich editorial description to facilitate discovery and use.
- 10.1.5 The result of this relative complexity is that the cost-per-hour of digitising audiovisual material is an order of magnitude higher than that for photographs, archival material and even complex 2D and 3D works of art, historical artefacts and specimen.
- 10.1.6 The process of establishing precise costings for the Digitisation of the legacy of analogue-format audiovisual material is complicated by the relative inaccuracy of available information about the number of hours of material to be digitised. Many collections are catalogued at collection-level, and most information about duration is based on estimates.
- 10.1.7 The successful Digitisation of audiovisual material is affected by the relative instability of current digital formats, many of which have a lifespan shorter than the original carrier material for the content. Hence, cultural institutions digitising audiovisual material must make a long-term commitment to the active curation of the material to ensure that it continues to be accessible in the long-term.
- 10.1.8 Because of the very significant costs involved, we have treated the Digitisation of audiovisual material separately from the collections held by museums, archives and libraries.

10.2 Scope of this enquiry

- 10.2.1 Audiovisual material exists in many different types of institution throughout Europe – from broadcasters to production companies, schools, universities, animation studios and many others.

10.2.2 For the sake of this enquiry, we will limit the scope to the film, video and audio material (including music and sound recordings) held in museums, archives and libraries.

10.3 Variables affecting AV Digitisation costs

10.3.1 As with all forms of Digitisation, there is no simple formula with which to calculate the costs of digitising AV material. The cost model includes the same elements as for other types of Digitisation workflow, and at each stage of the process, the associated costs will be affected by the organisational context, the fragility and complexity of the material, the scale of the Digitisation programme, the availability of specific expertise and a wide range of other interdependent factors.

10.3.2 The need to preserve the original transport medium during the Digitisation workflow, alongside the greater requirement for more complex metadata, means that workflow rates for audiovisual Digitisation are likely to be significantly slower than for formats that are more amenable to 'batch' or mass-processing.

10.3.3 Sampling rates and quality exert a particular influence on the costs of digitising AV material, as does the requirement for specific technical expertise at each state of the Digitisation process. Quality assurance is of particular importance to avoid compression artefacts such as 'ghosting' of video images arising from the conversion process.

10.3.4 Because of the relatively high initial capital costs of Digitisation equipment, skills and premises, the costs of digitising audiovisual material tend to respond to economies of scale – with lower costs attributable to outsourced Digitisation service providers or onsite programmes involving large quantities of material. Digitising audiovisual material on an on-demand or boutique basis tends not to scale in terms of cost.

10.3.5 For this reason, there are 3 possible models for establishing a Digitisation workflow for audiovisual material:

- a) Establishment of an in-house AV Digitisation facility, including the acquisition of equipment, staff and premises to support long-term in-house programmes;
- b) Outsourced (even when on-premises) Digitisation making use of equipment and skills owned by a 3rd party;
- c) Renting equipment and skills on a semi-permanent basis.

10.3.6 In many cases, it has proved more cost-effective in the long-term for cultural institutions to rent AV Digitisation equipment, since this protects them from the significant upfront capital outlay and allows them to spread expenditure over several years of the programme.

10.4 A Note about Rights

10.4.1 For the sake of simplicity, we have not addressed the full costs of retrospective rights clearance. For audiovisual material, this can be a complex and costly process, since video often includes layered rights, many of which may not be documented in any form.

10.4.2 Information about the costs and effort associated with rights clearance is available elsewhere in research commissioned by the European Commission.

10.5 Scale of AV Collections

10.5.1 In order to calculate the estimated costs of digitising AV material in museums, archives and libraries, we first need to establish a viable estimate for the quantity of material (in hours) they hold.

10.5.2 Turning first to the NUMERIC statistics produces the following estimates*:

<i>Institution type</i>	<i>Audio (average holdings in hours)</i>
Archive/Records Office	3520
Audiovisual or Film Institute	353572
Broadcasting Institute	253728
Art Museum	6317
Science/Technology Museum	679
Other Museum	886
National Library	168653
Higher Education Library	29463
Public Library	139724
Special or Other Library	6494

<i>Institution type</i>	<i>Film (average holdings in hours)</i>
Archive/Records Office	17021
Audiovisual or Film Institute	137245
Broadcasting Institute	26760
Art Museum	817
Science/Technology Museum	710
Other Museum	1131
National Library	58660
Higher Education Library	453
Public Library	2710
Special or Other Library	1676

<i>Institution type</i>	<i>Video (average holdings in hours)</i>
Archive/Records Office	2186
Audiovisual or Film Institute	38588
Broadcasting Institute	142767
Art Museum	254
Science/Technology Museum	272
Other Museum	574
National Library	78686
Higher Education Library	1946
Public Library	3200
Special or Other Library	192

* Note that all estimates are based on survey medians, not extrapolated figures

10.5.3 It is useful to compare these median figures against the findings of the comprehensive TAPE survey of audiovisual collections in European cultural institutions.

10.5.4 The TAPE²² survey produced the following estimates (based on the number of organisations that responded with details of their holdings of each material type):

²² TAPE survey of audiovisual material - <http://www.tape-online.net/survey.html>

<i>Format</i>	<i>Number of respondents</i>	<i>Total (x1000 hours)</i>	<i>Average (x 1000 hours)</i>
Audio	288	9386	33
Video	274	10559	39
Film	152	894	6
TOTALS	-	20839	-

10.5.5 These figures are significantly higher than the NUMERIC estimates – up to 9.4m hours of audio and over 10.5m hours of video. The TAPE survey further notes that a significant proportion of overall holdings are managed by a relatively small number of National Libraries and specialist Institutes.

10.5.6 The divisions across different institution types are illustrated, for example, by the following breakdown of Film holdings:

<i>Institution Type</i>	<i>Number of respondents</i>	<i>Total (x1000 hours)</i>	<i>% of total</i>
Archives	75	193	22
Libraries	16	34	3.9
Museums	22	60	6.7
Institutes	12	7.2	0.8
Research Institutes	7	1.7	0.2
Broadcasters	6	515	58
Commercial companies	5	65	7.3
Private Collections	3	0.8	0.1
Others	6	6.7	1.9
TOTALS	152	894	100

10.5.7 The TAPE figures are hampered (as are the NUMERIC statistics) by their relatively low response rate. However, since a significant majority of the total holdings of audiovisual material are held within a relatively small number of National institutions (for example, of the 9.4m hours of audio material, almost 5m hours are held in just 4 deposit and broadcast collections), it seems likely that the TAPE figures are more broadly accurate in respect of the overall holdings.

10.5.8 For the purposes of our estimation, therefore, we will extrapolate a figure based on the TAPE figures plus an adjustment of 15% to account for the smaller institutions that did not respond to or participate in the survey.

10.5.9 Taking this extrapolation provides the following estimates:

<i>Format</i>	<i>Total (hours)</i>	<i>Extrapolated (+15%)</i>
Audio	9.4m	10.81m
Video	10.56m	12.14m
Film	0.9m	1.04m
TOTALS	20.8m	29.92

10.5.10 Overall, then, we will base our calculation on a figure of approximately 29.92m hours of audio, video and film material, although in reality the total holdings of museums, archives and libraries are likely to be significantly greater than this.

10.6 Costs of Digitising Audio

10.6.1 The per-hour costs of digitising audio material will vary greatly depending on the nature of the material, the context within which it is taking place, the way in which the project is staffed and the output rate. An indicative set of high and low-end costings is provided below:

Low-end Audio Digitisation

<i>Item</i>	<i>Per-hour Cost (€)</i>
Project Management	3.50
Preparation & movement	4.50
Preservation	10.00
Conversion costs	25.00
Rich metadata enhancement	7.00
TOTAL per volume	€50.00 per hour

High-end Audio Digitisation

<i>Item</i>	<i>Per-hour Cost (€)</i>
Project Management	4.50
Preparation & movement	5.00
Preservation	13.00
Conversion costs	36.50
Rich metadata enhancement	15.00
TOTAL per volume	€74.00 per hour

10.6.2 Based on these estimated production costs, we can extrapolate max and min costs for digitising the audio collections mentioned in 10.5.9:

<i>Item</i>	<i>Total Cost (€bn)</i>
Minimum estimate (based on 10.81m hours @ €50.00 per hour)	0.54
Maximum estimate (based on 10.81m hours @ €74.00 per hour)	0.80
Mean	0.67bn

10.7 Costs of Digitising Video

10.7.1 Very little precise information is available about the costs of digitising video material, and in the course of this research, it has not been possible to identify precise or statistically-valid figures for Digitisation costs other than the specific conversion costs.

10.7.2 The original NUMERIC survey provides us with the following MEAN figures for video Digitisation, based on a relatively small survey sample:

<i>Institution Type</i>	<i>Number of respondents</i>	<i>MEAN cost per hour (€)</i>
Archives/Record Office	133	177.48
Audiovisual or Film Institute	31	-
Broadcasting Institute	10	55.18
Art Museum	224	417.70
Science Museum	25	-

Other Museum	83	64.58
National Library	30	35.88
HE/University Library	62	48.63
Public Library	65	-
Other Libraries	65	120.00
TOTALS	718	

10.7.3 These figures are problematic in that they do not indicate whether they are created on the basis of *solely* conversion costs, or an additional allocation of overhead. Given that all other NUMERIC figures are *solely* conversion rates, we propose to adjust these to account for a 30% equipment and 30% overhead apportionment. This is evidentially problematic, and the resulting calculations should be revisited in a future study when additional data is available.

10.7.4 Based on these figures, we can extract a minimum cost-per hour (median of a lower tercile) as €51.01 and a maximum (median of an upper tercile) as €238.39. Applying apportionment for equipment and overhead costs yields a cost-per-hour range of €153 to €715 – which is more in line with the evidence that it has been possible to collate from projects.

10.7.5 Based on these figures, we can extrapolate maximum and minimum figures for the costs of digitising the video material we estimate to be held in cultural institutions in the EU:

<i>Item</i>	<i>Total Cost (€bn)</i>
Minimum estimate (based on 12.14m hours @ €153.00 per hour)	1.86
Maximum estimate (based on 12.14m hours @ €715.00 per hour)	8.67
Mean	5.26bn

10.8 Costs of Digitising Film

10.8.1 As with the costs of digitising video, very little specific data is available concerning costings. The best estimates are provided by the PRESTOPrime project report²³, which states:

(p.10): "Film scanning equipment is expensive. One of the basic issues in creating the Presto project in 1999 was the high cost of film processes, as compared to videotape. As a rule of thumb, anything involving film would cost roughly ten times as much as a similar operation on videotape. Videotape copying and digitisation had a benchmark cost of €100 to €200 per hour, and film-to-film copying or film scanning/digitisation was indeed running at €1000 to €2000 per hour."

10.8.2 Based on these very broad estimates, we can calculate the following range for the cost of digitising our 1.04m hours of film material as:

²³ PRESTOPrime Audiovisual Digitisation Status Report (Wright, Richard), January 2010

<i>Item</i>	<i>Total Cost (€bn)</i>
Minimum estimate (based on 1.04m hours @ €1000.00 per hour)	1.04
Maximum estimate (based on 1.04m hours @ €2000.00 per hour)	2.08
Mean	1.56bn

10.9 Calculating the Totals

10.9.1 Based on the figures above, we can now calculate the total projected costs (albeit based on very loose projections) of digitising audiovisual material in European cultural institutions. In calculating these costs, however, we must be aware of the fact that respondents to the NUMERIC survey identified up to 34% of AV holdings as being incapable of Digitisation (largely due to the fragility of the medium).

10.9.2 Adjusting, therefore, for this proportion, we can derive the following estimated totals:

<i>Format</i>	<i>Quantity (m hours)</i>	<i>Total cost (unadjusted)</i>	<i>Adjusted total cost (66%)</i>
Audio	10.81	€0.67bn	€0.44bn
Video	12.14	€5.26bn	€3.47bn
Film	1.04	€1.56bn	€1.03bn
TOTALS	23.99	€7.49bn	€4.94bn

11. The Cost of Digitising European Cultural Heritage

11.1 Getting Started

11.1.1 We now have our estimated figures for the following 3 sets of variables:

- The number of cultural institutions of each type in Europe
- The approximate quantity of each type of material that they hold
- The approximate costs of digitising this material

11.1.2 We can, therefore, complete our calculation by adding together the total costs (shown in the table below) to reach a total estimated figure of €105.31bn.

<i>Description</i>	<i>Value</i>	<i>Units</i>
Estimated MEAN cost of digitising LIBRARY COLLECTIONS in the EU	19.77	€bn
Estimated total cost of digitising MUSEUM COLLECTIONS in the EU	38.73	€bn
Estimated total cost of digitising NATIONAL ARCHIVES in the EU	41.87	€bn
Estimated total cost of digitising AV COLLECTIONS in the EU	4.94	€bn
Estimated total cost of digitising CULTURAL MATERIAL in the EU*	105.31	€bn

11.1.3 This is not, however, quite the end of the story. Digitisation is a large-scale, multi-annual activity, and to digitise collections on the scale discussed in this document would take not only financial investment, but also a period of some 10-30 years.

11.1.4 The total cost should, therefore, be spread over this period. This is significant because we have already seen a significant reduction in per-unit Digitisation costs over the previous 10-year period, due to:

- Increased access to existing skills within cultural institutions
- Better and more efficient equipment and technology
- More accurate OCR and related software leading to a reduced error rate
- The emergence of Digitisation solutions that benefit from economies of scale and aggregation
- The emergence of a competitive commercial secondary market for Digitisation services

11.1.5 We must, therefore, assume that our calculation of the total cost of Digitisation would benefit from this 'Doppler-effect' in that it would cause costs to be reduced on an ongoing basis through the sheer investment in Digitisation practice and technologies, in effect reducing its own costs over the lifetime of the endeavour.

11.1.6 If we estimate an annual overall cost decrease of 0.5% over the past 10 years, based on comparison with historical cost data (which is admittedly partial), and assuming that this rate may continue at an equivalent level, we would envisage a cumulative saving of €5.14bn, resulting in a total cost of almost exactly €100bn over 10 years, or €10bn per annum.

11.2 The Cost of Ownership

11.2.1 The act of digitising a piece of material is only the beginning of a long-term relationship with the newly-created digital asset, and it is essential not to lose sight of the total lifetime costs of ownership, over and above the initial capital outlay on Digitisation.

- 11.2.2 Most estimates hold that the cost of ownership for a digital asset in an institutional context for 10 years will approximate 50-100% the cost of creating it in the first place.
- 11.2.3 In a context (such as an academic institution) where there is access to large-scale preservation infrastructure, this 10-year cost may drop to as little as 10-25% of the initial outlay on creation.
- 11.2.4 In the absence of a centrally-funded Digital Preservation infrastructure, then, we should be considering that the net expenditure on preserving the digitised record of European cultural heritage would be between €50bn and €100bn across the whole of the EU.
- 11.2.5 Where large-scale repository infrastructure is made available, this lifetime cost of ownership drops to approximately €10bn - €25bn.
- 11.2.6 At a net saving of between €40bn and €75bn over a 10-year period, these figures present a powerful argument for the coordinated and strategic large-scale investment in repository infrastructure for European cultural heritage as a net saving in the medium/long-term costs to the industry of maintaining the digitised materials.

12. Comparisons

12.1 Why compare?

12.1.1 In order to understand the implications of the proposed costs of digitising Europe's cultural heritage, it is useful to consider them in the context of other forms of public expenditure.

12.1.2 The proposed comparators for this study are:

- Total cost of development of the Joint Strike Fighter
- Cost of provision of library services in Europe
- Cost of building 100km of main road

12.2 The Joint Strike Fighter

12.2.1 The Joint Strike Fighter is a multi-nation joint procurement initiative to replace the current stock of fighter aircraft in use by the US Air Force, the UK Air Force and their various allies. The programme is being led by the US Air Force, which is also the majority funder.

12.2.2 The Research and Development costs of the Joint Strike Fighter have increased during the programme, but the current estimate stands at €40.34bn.

12.2.3 On completion, the purchase price for one Joint Strike Fighter aircraft is estimated to be around €147.41m. The annual maintenance costs are currently unknown.

12.2.4 The following figures give an indication of the comparative price of developing and buying the Joint Strike Fighter as compared to the Digitisation of cultural heritage. All cost estimates are based on mean averages.

- The purchase price of a Joint Strike Fighter is €147.41m, equivalent to:
 - Digitising 1m individual books if the majority of Digitisation is done in-house
 - Digitising 1.67m books if the Digitisation is outsourced
 - Digitising 2.42m books under a Public Private Partnership
 - Digitising 96,789 rare books, manuscripts and incunabula
 - Digitising 29.5m historic photographs
 - Digitising 1.83m man-made artefacts in museums
 - Digitising 2.02m natural artefacts in museums
 - Digitising 36.85m pages of archival records
 - Digitising 2.4m hours of audio material
 - Digitising 0.34m hours of video
 - Digitising 0.09m hours of film

12.3 The Provision of Library Services in Europe

12.3.1 In the absence of a valid overall figure for investment in public library services across the EU, we have taken the specific case of investment in UK libraries for 2008-09, which amounted to €1.72m.

12.3.2 This investment is equivalent to funding the digitisation of:

- 1129 rare books
- 344000 photos

- 21366 man-made artefacts
- 23562 natural history specimen
- 430000 pages of archival records
- 27742 hours of audio
- 3963 hours of video
- 1147 hours of film

12.4 The Cost of Building 100km of Main Road

12.4.1 Although the costs of building a length of road obviously vary considerably between nations, the average across Europe is estimated at between €500m and €1bn per 100km²⁴.

12.4.2 Taking an average of €750m to build 100km of main road, therefore, this investment is equivalent to funding the Digitisation of:

- 0.49m Rare books (7% of the total held in EU libraries)
- 150m photographs (40% of the total held in EU cultural institutions)
- 9.32m man-made artefacts (4% of the total in EU museums)
- 10.27m natural artefacts (5% of the total in EU museums)
- 187.50m pages of archival records (2% of the total holdings of EU National Archives)
- All 10.81m hours of audio material in EU cultural institutions
- 1.73m hours of video (14% of the total)
- 0.50m hours of film (48% of the total)
- 5.09m books digitised in-house (7% of the total)
- 8.49m books digitised in partnership with an external agency (11% of the total)
- 12.33m books digitised under a PPP (16% of the total)

²⁴ Source: <http://www.roadtraffic-technology.com>

13. Methodology

13.1 Accuracy and Scope

- 13.1.1 Digitisation is a set of business activities which collectively form a process. As with any other form of production or management process, the costing of these activities is subject to a wide range of variables. The nature of these variables is addressed in this report.
- 13.1.2 In order to establish a verifiable and justifiable cost model, we have drawn on published data concerning the costs of Digitisation projects at different scales, in different organisational contexts and in different European Countries. The cost estimates given in this report are given with an upper and lower tolerance, and should in all cases be adjusted to the specific context of a particular country, organisation or type of material.
- 13.1.3 The end result of this process is a set of ranges which should be regarded as broadly indicative, rather than specifically accurate. As stated at the outset of this exercise, the variables involved in providing specific costs are too many, too interdependent and too dependent on the specific organisational context to lend themselves to a completely normalised approach.

13.2 NUMERIC Data

- 13.2.1 This report has depended heavily on the data provided by the EU-funded NUMERIC project, as the latest, most recent, most comprehensive and most reliable source of statistical data on Digitisation and associated costs.
- 13.2.2 However, the validity of the NUMERIC data has been contested on the basis of its identification of a target group of 'relevant' institutions, and this does indeed mean that the data must be regarded as indicative rather than actual.
- 13.2.3 Wherever possible, findings based on the NUMERIC dataset have been contrasted with actual costs taken from current Digitisation projects or other research materials.

13.3 Source Data

- 13.3.1 In the development of this study, National organisations and funding agencies were contacted to provide data and evidence concerning actual expenditure on Digitisation in national projects.
- 13.3.2 A considerable quantity of source data was gathered from existing published data about project costs.
- 13.3.3 Data on costs has also been provided by commercial Digitisation services.

13.4 Analysis

- 13.4.1 Analysis of the resulting data has been conducted on the basis of a series of reasoned calculations, adjusted and weighted depending on the reliability of the available data and the resulting extrapolations.

14. Bibliography

14.1 Sources

14.1.1 In the course of this study, we have made reference to the following sources:

<i>Title</i>	<i>Author (where known)</i>	<i>Publisher & date</i>
Assessment of the Orphan Works issue and the costs for Rights Clearance	Vuopala, Anna	European Commission, May 2010
Cost Models for Digitisation (PrestoSpace report)	Addis, Matthew	PrestoSpace, July 2005
Digitisation Costs & Funding Workshop	N/A	OCLC, Oct 2003
eIFL Case Studies on Low Cost Digitisation Projects	de Vries, Repke	eIFL, Oct 2009
Final Report on Digital Preservation, Orphan Works & Out-of-print Works	HLEG	European Commission, June 2008
IFLA Conference Proceedings on Digitisation	N/A	IFLA, Aug 2009
INSAR Newsletter (issue 10)	N/A	Jun, 2005
JISC Digitisation Project Case Studies (various)	N/A	JISC, 2008-2009
Life3: Predicting Long-term Digital Preservation Costs	Wheatley, Paul & Hole, Brian	British Library, 2009
NUMERIC Report	N/A	IPF, 2009
NUMERIC Report Appendices	N/A	IPF, 2009
The Electronic Archive Project in Hungary	Szaticsek, Zoltan	MOL, April 2009
The National Archives Commercial Digitisation Programme	N/A	SCA, 2009
The Price of Digitisation: Report of the 2003 NINCH Symposium	Hughes, Lorna M.	NINCH, April 2003
The Status of Digitisation in Europe	Ayris, Paul	LIBER/EBLIDA, Feb 2010
Understanding the Cost of Digitisation	Hammond, Max & Davies, Claire	Curtis & Cartwright, Oct 2009
LIFE: Costing the Digital Preservation Lifecycle	Wheatley, Paul & Ayris, Paul	British Library
Estimating Digitisation Costs in Digital Libraries using DiCoMo	Bia, Alejandro, Munoz, Rafael & Gomez Jaime	Springer, 2010
Tracking the Reel World, a survey of Audiovisual Collections in Europe	Klijn, Edwin & de Lusenet, Yola	European Commission, 2008

14.1.2 In the course of this study, we have also made reference to the following online sites/services (starts overleaf:

Akademie der bildenden Künste Wien	http://www.akbild.ac.at/kuka/
Association of Moving Image Archivists	http://www.amianet.org/
Bibliothèque Nationale de France	http://www.bnf.fr
Centre National de l'Audiovisuel (CNA)	http://www.cna.public.lu/
Centre Virtuel de la Connaissance sur l'Europe (CVCE)	http://www.ena.lu
Council of Europe	http://consilium.europa.eu
Culturenet Denmark	http://www.kulturnet.dk/en/text/
CultuurDatabank Vlaanderen	http://www.cultuurnet.be/cultuurdatabank
Danish Film Institute	http://www.dfi.dk
Open Content Alliance	http://www.opencontentalliance.org
ETHOS Toolkit	http://ethostoolkit.cranfield.ac.uk
European Commission Audiovisual Services	http://ec.europa.eu/avservices/index.cfm?sitelang=en
European Film Gateway	http://www.europeanfilmgateway.eu/
EuropeanaConnect	http://www.europeanaconnect.eu/
EUScreen Project	http://www.euscreen.eu/
Finnish Museum of Photography	http://www.fmp.fi
InfoSoc Directorate, European Commission	http://ec.europa.eu/dgs/information_society/index_en.htm
International Association of Sound and Audiovisual Archives	http://www.iasa-web.org/
International Federation of TV Archives	http://www.fiatifta.org
Joanneum Research (Graz)	http://iis.joanneum.ac.at/iis/Default.asp
Kunsthistorisches Museum	http://www.khm.at/
LIFE Project Website	http://www.life.ac.uk
Maerlant Center	http://www.maerlant.be
National Cultural Heritage Agency	http://www.kuas.dk
Netherlands Institute for Image and Sound	http://www.nibg.nl
NUMERIC WS	http://www.numeric.ws/
PrestoPRIME	http://www.prestoprime.org/
PrestoSpace	http://prestospace.org/
UNESCO Audiovisual Archives	http://www.unesco.org
UNESCO Communication and Information	http://www.unesco.org
VideoActive	http://www.videoactive.eu/VideoActive/Home.do