THE NEW RENAISSANCE

REPORT OF THE ‘COMITÉ DES SAGES’

REFLECTION GROUP ON BRINGING EUROPE’S CULTURAL HERITAGE ONLINE
A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu).

Cataloguing data can be found at the end of this publication.


doi:10.2759/45571

Reproduction is authorised provided the source is acknowledged.

Printed in Luxembourg

PRINTED ON ELEMENTAL CHLORINE-FREE BLEACHED PAPER (ECF)
The new Renaissance

REPORT OF THE ‘COMITÉ DES SAGES’
ON BRINGING EUROPE’S CULTURAL HERITAGE ONLINE

Elisabeth Niggemann
Jacques De Decker
Maurice Lévy
Foreword

There is probably no greater ambition than to perpetuate our rich cultural heritage. It is therefore in full consciousness of our responsibility towards past and future generations and in deep humility that we have approached our mission.

As Jean Monnet said, if ‘Europe were to be reconstructed, I would begin with culture rather than the economy’. The cultural heritage of the old continent nourished the education, the formation, the spirit of the generations which preceded us and we feel the responsibility to transmit this rich (indeed, one of the richest in the world) heritage to future generations and to make sure it will be preserved, enriched and shared.

With no exaggeration, we can state that what is at stake is a common good of humanity and not just of Europe.

We therefore looked for solutions which would allow:

- To protect this invaluable asset that is our culture, the heritage of which we are the guardians and which it is our duty to pass on.
- To make it accessible to the greatest number of people without distinction or barrier.
- To ensure that this inheritance can remain a living asset over time and that it is as widely shared as possible.
- To make sure that the creators and all those working to produce and broadcast their work can enjoy the fruits of their labours and that creativity can blossom without hindrance.
- Not to impose selection for protection and preservation. On what grounds would we have the right to lay down selection criteria for what should or should not deserve to be protected? We did not feel we had the right to set criteria for selection.
- To ensure that the financing meets the fundamental principle of accessibility for all (hence the necessity of requesting public funding), but also the reality which confronts us today (the acute lack of public resources, not to mention sovereign debts). And consequently the necessity of defining what could be the guidelines for public-private partnerships, and other forms of financing.
To envisage the commercial, economic or growth opportunities which digitisation could produce for Europe and how it could generate jobs.

We have conducted our mission with the greatest respect for the works, the authors, the producers/publishers and the public. We have sought to promote the most open-minded approach towards all parties involved and, above all, the greatest ambition for our rich heritage.

We were proud to undertake this mission. Unfortunately we had to finish our report by a set date, and we are conscious that, while it does address some of the key issues mentioned in the mandate given to us by Commissioners Kroes and Vassiliou, we could have worked much longer and dug deeper to cover some of the thorny challenges that the evolution of technology poses: authors’ rights versus copyrights, user-generated content, downloads, etc.

We would like to express our thanks to Commissioners Kroes and Vassiliou and to the European Commission for their confidence, with the hope that this report, which reflects the view of the Comité, will help the European Union and Member States to further define policy in these fields.

We would also like to express our thanks to the experts of Cap Gemini who have worked pro bono for the Comité des Sages. They have assessed the technology used by Europeana to see whether it can be considered as ‘future-proof’. Their point of view has helped the Comité to determine the most effective recommendations for the next steps for Europeana.

Finally we would like to warmly thank the following persons who helped us navigate through the complexity of the issues and provided us with invaluable support: Yvo Volman, Anna Athanasopoulou and Valérie Panis, working for the European Commission, Britta Woldering of the Deutsche Nationalbibliothek and Maxime Baffert working for Publicis Groupe.

Elisabeth
Niggemann

Jacques
De Decker

Maurice
Lévy

Brussels, 10 January 2011
Table of contents

Foreword .................................................................................................................. 5

Executive Summary .................................................................................................. 8

Chapter One: The Mission ...................................................................................... 13

Chapter Two: The Vision ......................................................................................... 14

Chapter Three: The Process ..................................................................................... 16

Chapter Four: Access and Use Models for Digitised Public Domain Material .......... 18

Chapter Five: Digitisation and Online Accessibility of In-Copyright Material .......... 22

Chapter Six: Europeana ............................................................................................ 28

Chapter Seven: Sustainability of Europe’s Cultural Heritage Online ......................... 32

Chapter Eight: Funding Digitisation and Europeana, Costs and Opportunities .......... 38

Chapter Nine: Public-Private Partnerships for Digitisation Opportunities, Challenges and Pitfalls .......................................................... 45

Chapter Ten: Conclusion .......................................................................................... 50

Annexes .................................................................................................................... 53
or centuries, libraries, archives and museums from across Europe have been the custodians of our rich and diverse cultural heritage. They have preserved and provided access to the testimonies of knowledge, beauty and imagination, such as sculptures, paintings, music and literature. The new information technologies have created unbelievable opportunities to make this common heritage more accessible for all. Culture is following the digital path and ‘memory institutions’ are adapting the way in which they communicate with their public.

Digitisation breathes new life into material from the past, and turns it into a formidable asset for the individual user and an important building block of the digital economy.

We are of the opinion that the public sector has the primary responsibility for making our cultural heritage accessible and preserving it for future generations. This responsibility for and control over Europe’s heritage cannot be left to one or a few market players, although we strongly encourage the idea of bringing more private investments and companies into the digitisation arena through a fair and balanced partnership.

Digitising our cultural heritage is a gigantic task that requires large investments. According to a study, in total some €100bn will be necessary over time to bring our complete heritage online. This type of effort needs time and the investment will need to be carefully planned and co-ordinated in order to get the best results.

We think that the benefits are worth the effort. These benefits are in the first place related to the wider access to and democratisation of culture and knowledge, as well as the benefits for the educational system - both schools and universities. Other major benefits lie in the economic sphere and concern the development of new technologies and services for digitisation, for digital preservation and for interacting in innovative ways with the cultural material. The digitised material can in itself be a driver of innovation and be at the basis of new services in sectors such as tourism and learning.

We make our recommendations with these potential benefits in mind and with the aim to promote an environment that will help to:

- share our rich and diverse common heritage
- link the past with the present
- preserve this heritage for future generations
- protect the interests of European creators
- nurture creativity, including creative efforts by non-professionals
- contribute to education, and
- spur innovation and entrepreneurship.

The recommendations concern all the areas identified by the terms of reference for our work, and address situations where we think
that a stimulus is necessary or barriers need to be removed.

1. **Ensuring wide access to and use of digitised public domain material**

   - Cultural institutions should make public domain material digitised with public funding as widely available as possible for access and re-use. This cross-border access should be part of the funding conditions for digitisation across Europe. The use of intrusive watermarks or other means that limit the use of the material should be avoided.

   - Where cultural institutions charge private companies for the re-use of the digitised public domain material, they should **comply with the rules of the European Directive on the re-use of public sector information**.

   - The European Commission should consider ways and means to **eliminate the differences in the rights status of digitised material between the Member States** in a context where cross-border access and use is the norm. In principle the mere digitisation process should not generate any new rights.

   - **Metadata** related to the digitised objects produced by the cultural institutions should be widely and freely available for re-use.

2. **Stimulating the digitisation and online accessibility of in-copyright material**

   - A **European legal instrument for orphan works** needs to be adopted as soon as possible. The instrument should **comply with the 8-step-test** as defined by the Comité.

   - **Future orphan works must be avoided**. Some form of registration should be considered as a precondition for a full exercise of rights. A discussion on adapting the Berne Convention on this point in order to make it fit for the digital age should be taken up in the context of WIPO and promoted by the European Commission.

   - National governments and the European Commission should promote solutions for the **digitisation of and cross-border access to out of distribution works**.

   - Rights holders should be the first to exploit out of distribution works.

   - For cultural institutions **collective licensing solutions and a window of opportunity** should be backed by legislation, to digitise and bring out of distribution works online, if rights holders and commercial providers do not do so.

   - Solutions for orphan works and out of distribution works must **cover all the different sectors**: audiovisual, text, visual arts, sound.

3. **Reinforcing Europeana as the reference point for European culture online**

   - Europeana should be further developed to become the reference point for European cultural content online. This requires a **concentration of financial efforts and political capital** at European
and at the national level for the development of the Europeana site and the underlying structures.

- Member States should ensure that all public funding for digitisation is conditional on the subsequent free accessibility of the digitised material through Europeana. They should also ensure that, by 2016, they have brought all their public domain masterpieces into Europeana.

- In the coming few years, Europeana should add to its portal an application platform, and main activities related to the digitisation and preservation of Europe's cultural heritage should be linked to the site. In the technical development of the site particular attention should be paid to multilingual aspects. Europeana should also explore the opportunities of cloud computing in the future.

- For the medium term, it should be considered to give Europeana a key role in the preservation of Europe's heritage and to turn it into a European deposit site for public domain digitised cultural material and into a dark archive for born digital cultural material.

- Europeana must be actively and widely promoted by the cultural institutions, by the European Commission and by the Member States.

4. Guaranteeing the sustainability of digitised resources

- Preservation is a key aspect in digitisation efforts. Digital preservation is also a core problem for any born digital content. The organisational, legal, technical, and financial dimensions of long term preservation of digitised and born digital material should be given due attention.

- The preservation of digitised and born digital cultural material should be the responsibility of cultural institutions - as it is now for non-digital material.

- To guarantee the preservation of the European digital cultural heritage, a copy of digitised or born digital cultural material should be archived at Europeana. For in-copyright works the deposit site would be a dark archive functioning as a safe harbour.

- To avoid duplication of effort by companies operating across borders and by the cultural institutions, a system could be envisaged by which any material that now needs to be deposited in several countries would only be deposited once. This system would include a workflow for passing on the copy to any institution that has a right to it under national deposit legislation.

- Copyright and related legislation have to enable the cultural heritage institutions responsible for preservation to create archival copies and to make file conversions for archival purposes.

---

1 A dark archive is an archive to which access is restricted.
• **Persistent identifiers** must be implemented in each digital object archived in cultural institutions. A reliable resolution service for **persistent identification** of digital objects must be developed and maintained on European level, preferably linked to Europeana.

5. **Finding sustainable financing for digitisation and Europeana**

• The **public sector has the primary responsibility to fund digitisation**, and Member States will need to considerably step up their investments in digitisation. The current financial crisis cannot be ignored, but equally cannot be a reason for not acting.

• The **involvement of private partners should be encouraged**. Private funding for digitisation is a **complement to the necessary public investments** and should not be seen as a substitute for public funding.

• **Digitisation** should in principle be funded at the **national or regional level, not at the European level**. However, the Member States should be strongly encouraged to use the funding possibilities from the European Structural Funds for digitisation activities. Also, some targeted digitisation efforts with a clear cross-border scope (e.g. cross-border collections) could be co-funded at European level.

• Given Europeana's character as a common good, **public funding should cover the largest part of Europeana's operational costs**, also after 2013. The funding of digitisation and of Europeana should be seen as a package, where **MS are broadly responsible for funding the digitisation** of their cultural heritage and creating national aggregators and where the **funding of the Europeana portal should come predominantly from the budget of the European Union**.

• Member States should **promote ways to turn digitisation into new development opportunities for European firms**, for example through regional clusters of businesses in partnership with cultural institutions, knowledge partnerships between cultural institutions and universities, or through strategic partnerships at European or international level in the area of new technologies and applications in relation to cultural heritage.

6. **Complementing public funding through public private partnerships for digitisation**

• In order to protect the interests of public institutions entering into a partnership with a private partner a set of **minimum conditions** should be respected:
  - The **contents of the agreement** between a public cultural institution and a private partner should be made public.
  - The **digitised public domain material should be free of charge for the general public and available in all EU Member States**.
  - The private partner should provide cultural institutions with **digitised files of the same quality** as the ones it uses itself.
The **maximum time of preferential use** of material digitised in public-private partnerships **must not exceed 7 years**. This period is considered adequate to generate, on one hand, incentives for private investment in mass-digitisation of cultural assets, and, on the other, to allow sufficient control of the public institutions over their digitised material.

Policy makers at European and national level should **create favourable conditions** for the involvement of European players. In particular:

- **Encourage digitisation in new areas** that have not received much attention thus far, such as audiovisual material, newspapers, periodicals or museum objects.
- In the medium term, subject to an improvement of the financial situation in the Member States, **create incentives for the investment of private funds through taxation**.
- **Encourage the use of public funds** matching private funds invested in digitisation. Public funds may be given to cultural institutions which have secured a partnership for the digitisation of their collection with a private entity, on a matching basis with the private funds invested.
- **Encourage Europeana** and its contributing institutions to **expand their digital contents by building partnerships with European businesses**.

Can Europe afford to be inactive and wait, or leave it to one or more private players to digitise our common cultural heritage? Our answer is a resounding ‘no’. Member States, Europe’s cultural institutions, the European Commission, and other stakeholders will all have to take up their responsibilities in order to ensure that Europe’s citizens and economy fully benefit from the potential of bringing Europe’s cultural heritage online.

Our goal is to ensure that Europe experiences a digital Renaissance instead of entering into a digital Dark Age.
In April 2010, Neelie Kroes, Vice President of the European Commission for the Digital Agenda, and Androulla Vassiliou, European Commissioner for Education, Culture, Multilingualism and Youth, appointed the Comité des Sages (Reflection Group) on bringing Europe’s cultural heritage online. The creation of the Comité followed a suggestion made by the French Minister of Culture and Communication during the debate on digitisation policies in the Education, Youth and Culture Council meeting of 27 November 2009.

This initiative was prompted by the need to reconcile the traditional mandate of cultural institutions, which give access to our common heritage and preserve it for future generations, with the opportunities and challenges emerging in the digital era. Managing the ‘digital shift’ requires a strategic approach that will ultimately define how access and use of digitised cultural resources can be facilitated in the European Union, and, eventually, how European creativity will be preserved and promoted for future generations.

These considerations lie at the core of the mandate guiding the work of the Comité.

The Comité was invited to provide a set of recommendations for the digitisation, online accessibility and preservation of Europe’s cultural heritage in the digital age, looking in particular at the issue of public-private partnerships for digitisation in Europe.

Its mandate covered in particular the following areas:

- the overall financial cost and level of public funding available to European cultural institutions for digitising their collections in Europe;
- the best models to maximise access and use of digitised material for the economy and society at large, in particular the fundamental conditions that should be respected in public-private partnerships for digitisation of works that are in the public domain;
- the role and responsibilities of private and public organisations for digitising orphan works as well as material that is in-copyright but no longer commercially available, with a view to tackling the risk of a ‘20th century black hole’ in Europeana and on the Internet in general;
- promoting the widest access to the digitised material across borders;
- ensuring sustainability of the digitised resources for long term preservation purposes.

The members of the Comité were asked to conduct their mission under their sole responsibility and to report back to the Commissioners within 8 months after their first meeting.
If one word should encompass and summarise the vision of the Comité des Sages, it would be ‘access’. When it comes to our common cultural heritage, there is no bigger challenge, there is no more urgent question than to secure the access of current and future generations to this heritage. Access to the largest population, both European and non-European. And access to one of the richest cultural heritages in the world, a universal common good.

Giving access to the European works and productions must become the yardstick of all initiatives taken in this field. And the many questions that have arisen with digitisation should be analysed through this prism and solved with the objective of making our heritage increasingly accessible to all. For instance, the proposals of the Comité regarding Europeana must be understood as a way to increase access to digitised goods, as Europeana is today one of the common tools of the European Union to serve this goal.

The principle that everything must be done to guarantee access to our cultural heritage has led to three other principles, which have guided the Comité in its work.

1. First, our cultural heritage is not only the legacy of the past, but is a body of knowledge, imagination and creativity which is constantly evolving and growing every day. Today’s wealth of cultural expressions and knowledge will be our common cultural heritage tomorrow. Therefore, although the recommendations focus on digitisation and the cultural heritage from the past, they always include what is added in the present because the past and the present must be available to future generations. One of our core missions is to ensure full access to cultural expressions and knowledge of the past, the present and the future.

2. Second, digitisation is more than a technical option, it is a moral obligation. In a time when more and more cultural goods are consumed online, when screens and digital devices are becoming ubiquitous, it is crucial to bring culture online (and, in fact, a large part of it is already there).

If we don’t pursue this task, we run the risk of progressively eroding and losing what has been the foundation of European countries and civilization in the last centuries. It must be clearly understood that if access is the final objective, a tall order, it can only be achieved through preservation of the work.

3. Third, the Comité believes that the main responsibilities for digitising and preserving our cultural heritage should be assumed by the public sector. In fact, it is the task of public authorities and cultural institutions to make sure that our heritage is properly digitised and then preserved in order to guarantee access to the largest
possible audience today and tomorrow. This is too important a responsibility to lie solely in the hands of private interests, which do not necessarily have the same long-term approach and public interest as governments. As culture is Europe’s most important common good, governments and their cultural institutions have the duty to be at the forefront of the digitisation effort.

Yet, two caveats should immediately be added to these principles.

1. First, even if the Comité des Sages is firmly convinced that digitisation is mainly a task for public authorities, this does not mean that the private sector should not be involved. Going back to our main objective to enlarge access to our cultural heritage, partnering with the private sector can be a good option to speed up the digitisation of works and collections from the public domain, at a time when public funding is scarce. Moreover, digitisation can also be taken as an opportunity to nurture and develop a whole new range of private activities and initiatives, through the new technologies and processes that have been (and will be) developed, and contribute to job creation in Europe.

2. Second, the fact that cultural goods should be accessible on line does not mean they should necessarily be available for free. Digitisation should not be a pretext to downgrade the protection that has been granted to authors and creators in most European countries. In fact, digitisation and protection of the creators’ rights are more intertwined than most people think: digitisation protects the artists from the past, while authors’ rights and copyright protect the artists of today and tomorrow. Hence, when works are still in copyright, it is obvious that authors and creators are compensated for their talent and their effort when their work is brought online. Even in the cases of orphan works or out-of-distribution works, the ‘imperative of access’ does not mean that all protection and compensation for the creators should be forgone. On the contrary, digitisation represents a new opportunity to give a new life to these works.

The Comité worked on the understanding that culture is not only limited to the traditional field of arts and letters, but also encompasses science and the entire body of knowledge and experience which our societies have been consolidating over the centuries. This is our heritage and our culture which we need to preserve and make accessible to all in the digital era.

This is the vision, the framework that the Comité has used to answer the questions raised in its terms of reference. The Comité believes that the implementation of these principles and of the recommendations of the report can help European countries to transform a challenge – how to avoid the disappearance of our heritage – into an opportunity for cultural development, artistic creation and economic growth.
The work of the Comité started after its official establishment on 23 April 2010 and was structured around a series of 8 working meetings, which were complemented by many interactions with stakeholders. In order to fulfill its mission, the Comité has organised an inclusive process, widely consulting the stakeholders and the general public.

Between 18 August and 30 September 2010 the Comité held a public consultation through the website of the European Commission, covering all the aspects of its mandate. The consultation yielded 1258 replies from 42 different countries. More than half of the replies came from private citizens, which shows the broad interest in the topic. 17% of the replies came from private companies or associations of companies and 14% from cultural institutions. An overview of the replies to the consultation is annexed to this report.

A public hearing took place in Brussels on 28 October and was attended by more than 100 stakeholders. 34 organisations submitted position papers in the context of the hearing, while 23 of them intervened during the session and exchanged views with the members of the Comité.

On 18 November 2010, the members of the Comité had a working lunch with the EU Ministers responsible for culture and audiovisual matters in the context of the Education, Youth and Culture Council. The same day the Comité had an exchange of views with the Culture and Education Committee of the European Parliament.

Furthermore, the members of the Comité had a series of formal and informal meetings with stakeholders either collectively or individually (authors and creators, publishers, technology firms, public authorities).

The Comité commissioned two external assignments to underpin its work. The first assignment consisted of a study on the cost of digitising Europe's cultural collections. The second assignment was a technical audit of the Europeana site. In the course of its discussions the Comité has also drawn upon earlier analytical work conducted at EU level and in the different Member States, upon the final report of the High Level Group on Digital Libraries of December 2009, as well as the reports submitted to the Commission by all the Member States on the digitisation, online availability and preservation of their cultural heritage.

Furthermore, the work of the Comité has been informed by a number of recent po-

---


2 The position papers, a video of the hearing and summary minutes can be found at the following web-page: http://ec.europa.eu/information_society/activities/digital_libraries/comite_des_sages/

3 The study results are annexed to this report.

4 The audit results are annexed to this report.
Political documents such as the resolution of the European Parliament on ‘Europeana, the next steps’ (Trupel report) adopted on 5 May 2010, the Council Conclusions on ‘Europeana, next steps’ adopted on 10 May 2010, and the Commission’s Communications on ‘Europeana next steps’ (28-8-2009), ‘Copyright in the knowledge economy’ (19-10-2009), and ‘A Digital Agenda for Europe’ (29-5-2010).
The issue

Cultural institutions across Europe have started digitising their collections of books, newspapers, maps, sound and video material, manuscripts, photographs and museum objects and making them available online. Large parts of the digitised works are in the public domain1 - that is no longer covered by intellectual property rights. They can therefore in principle be accessed and used for free by everyone.

The present chapter deals with material digitised with public money, coming from the budgets of the cultural institutions or from digitisation projects funded at national, regional or local level. The case of public private partnerships for digitisation is treated in chapter 8 of this report.

The cultural sector has adopted a wide range of practices in giving access to or allowing the use of public domain material digitised with public money. Many of the institutions make the material freely available online, but some charge for downloading and sometimes for accessing digitised items.

In practice, the limitations on the use of the material are determined by the use condi-

---

1. Several institutions clearly indicate that material is in the public domain. They allow free access and download by non-commercial users. Some of them charge for the commercial re-use of the material. When users download material, they are asked to confirm that they will use the material only for non-commercial purposes.

2. Some institutions charge for accessing their collections on a subscription basis, including for accessing the public domain material. One institution charges for viewing the historical material on the basis of a subscription for a few hours or for a longer period.

3. There are also examples where the user is allowed to see the full content
on the screen, but a fee is charged for downloads of the public domain material.

Several contributing institutions add intrusive watermarks to their digitised public domain objects, to avoid unauthorised re-use. This practice is for example followed for old manuscripts and incunabula and for images. The problem is that this type of watermarks makes the material unusable, even for schoolwork.

Some of the cultural institutions, in particular museums, only make images available in a low resolution (normally good enough for private use, e.g. schoolwork). They charge for the access to and use of the same images in a higher resolution.

METADATA
Cultural institutions add considerable amounts of information to digitised objects (metadata), describing for example the author, the provenance and age of the work, giving contextual information, as well as technical information on the formats used and characteristics allowing search engines to locate the object. This metadata is essential to provide the user with a useful background to the work, and also to allow search-engines such as Europeana to locate the digitised objects relevant in the context of a specific search.

In some cases cultural institutions charge for or impose other conditions for the re-use of this metadata and they tend to be particularly wary of the commercial use of the data. This commercial use is broadly defined, and can for example include the indexing by commercial search engines. The conditions imposed on the use of the metadata can in turn pose problems for services such as Europeana, which want to open up the metadata they harvest in order to get the broadest possible exposure.

General considerations

The Comité feels that, in order to tackle the issue of access to and use of digitised public domain material, its work should draw in the first place upon basic principles of access to culture. People have the right to access culture and knowledge, and wide access and use also serve the public interest to preserve cultural diversity.

At the same time one should take into account the reality in which cultural institutions operate (‘principe de réalité’), includ-
ing financial constraints, which are likely to become more stringent in the coming years.

Differences in the rights status of digitised material between the Member States should be eliminated in a context where cross-border access and use is the norm. In principle the mere digitisation process should not generate any new rights.

**Access to the digitised material**

The Comité feels strongly that public domain material digitised with public money should be freely accessible for all. This should be part of the funding conditions for the digitisation of public domain material across Europe.

This is a matter of principle and also rooted in practical considerations. Google has set a standard of free access for the end-user to major collections of digitised public domain books through Google Books. Models in which cultural institutions charge the end-user for accessing the public domain resources they have digitised are therefore unlikely to be sustainable, since non-competitive in market terms.

If cultural organisations have to generate part of their operating costs by monetising their digitised public domain content, they should consider alternative revenue streams - such as sponsoring, partnerships, advertising and/or services related to the use of cultural material - rather than charging for online access to the material.

The Comité considered the question of whether restrictive practices by cultural institutions should have consequences, for example the exclusion of the material from Europeana. The Comité is of the opinion that material from cultural organisations which charge for accessing the public domain material should not be excluded from the site, since Europeana has the vocation of giving access to the widest possible range of digitised cultural material. There should, however, be a possibility for the user to filter searches in Europeana in order to obtain only material that can be accessed for free.

At the same time cultural institutions should be encouraged to comply with the Europeana public domain charter and make their digitised public domain content available for free.

**Re-use of the digitised material**

Public domain material digitised with public money should be freely available for non-commercial re-use by citizens, schools, universities, non-governmental and other organisations.

The Comité also recommends that cultural institutions make their digitised material available as widely as possible for commercial entities to build upon. This will stimulate the use of the material for new information products and services.

Cultural institutions can, however, ask private companies to pay for the commercial re-use of the material, in order to recoup the digitisation costs and finance further digitisation. This could be done on the basis of a one-off payment or through revenue-sharing models. Specific partnerships between cul-
tural institutions and smaller companies or universities could be considered, in order to stimulate the re-use of the digitised material for innovative information services.

This should not lead to exclusive arrangements or market discrimination. In general, cultural institutions allowing the re-use of their digitised material should comply with the rules of the EU Directive on the re-use of public sector information. This Directive regulates the behaviour of public sector bodies when they influence the information market by selling information or making it available for re-use. The Directive builds on two of the pillars of the internal market, transparency and fair competition. Currently, cultural institutions are excluded from the scope of the Directive.

The Comité considers the use of intrusive watermarks on digitised objects to be a particularly bad practice, and cultural institutions using this type of watermark should be asked to reconsider their policy. This does not apply to smaller watermarks at the bottom of a digital object indicating the origin of the material, which do not present any problem.

Use of metadata

The Comité is of the opinion that metadata related to the digitised objects produced by the cultural institutions should be widely and freely available for re-use. If conditions are imposed (e.g. a reference to the producer of the metadata), these should be standardised across the board.

Key Recommendations

1. Cultural institutions should make public domain material digitised with public funding as widely available as possible for access and re-use. This should be part of the funding conditions for digitisation across Europe. The use of intrusive watermarks or other means that limit the use of the material should be avoided.

2. Where cultural institutions charge private companies for the re-use of the digitised public domain material, they should comply with the rules of Directive 2003/98/EC on the re-use of public sector information.

3. The European Commission should consider ways and means to eliminate the differences in the rights status of digitised material between the Member States in a context where cross-border access and use is the norm. In principle the mere digitisation process should not generate any new rights.

4. Metadata related to the digitised objects produced by the cultural institutions should be widely and freely available for re-use.
Digitisation and online accessibility of in-copyright material

The issue

Copyright covers all forms of expression - ranging from personal letters to commercially produced material such as films, music and books as well as radio and television broadcasts produced by private companies and public sector broadcasters. All of these types of material are held by cultural institutions in Europe.

The ease with which today's users can access big search engines and platforms and find an overwhelming offer of information, books, newspapers, websites, archival material, pictures, music or movies naturally leads to expectations towards cultural institutions. Accustomed to the comfort of search engines and new services, they expect to find everything on the web. ‘What is not on the web, does not exist’ is the core of their belief and behaviour. What is on the shelves, in the archives, in the exhibition halls of cultural institutions will soon fall into oblivion, if it is not digitised and offered alongside the born digital works and all the other internet services.

Therefore, cultural institutions have been trying to digitise as much of their holdings as possible. For financial and practical reasons, they started by digitising their out-of-copyright holdings. Obviously, before digitising in-copyright works and offering them on the internet, rights have to be cleared. This is time consuming and extremely costly. As an indication, at the public hearing organised by the Comité, the German broadcaster ZDF estimated the number of contracts related to its archives at 3 million - something that makes individual renegotiation practically impossible - and mentioned that the BBC has calculated that clearing rights for the whole BBC archive would cost 72 million pounds for staff costs alone.¹

Today’s publishers, music, film and TV producers have been creating attractive offers for their born digital material. What is missing, though, is the in-copyright-material, that is not on offer by the respective distributors and right owners. There is a „black hole of the 20th century“, in which the majority of the traditional works of the last century falls, because they are not digital, they are out of distribution and quite often they are orphan works.

¹ http://ec.europa.eu/information_society/activities/digital_libraries/comite_des_sages/
Orphan works, works where the rights holder cannot be identified or located, form a barrier to mass digitisation projects. Orphan works exist in all domains, text, music, film etc and increase in number the further back in time one goes, or the more complex the layers of rights become. The Association des Cinémathèques Européennes estimates that 21% of films held in audiovisual archives are orphaned, with 60% of these being over 60 years old. The British Library estimates that 40% of its in-copyright collections are orphan. The ‘In from the Cold’ report estimated approximately 90% of the photographic record in UK cultural institutions as orphaned.

There are quite a number of studies from the cultural sector highlighting the issues of rights clearance focussing specifically on the issue of orphan works in the field of books, audio recordings and audiovisual material. As the High Level Expert Group on Digital Libraries recommended in 2008 there is a need to develop legal solutions to adequately address this issue in order to create legal as well as reputational certainty for the digitising partners.

Europeana has been founded to be the platform for Europe's cultural heritage. But a European digital library, archive and museum which lacks the works of the 20th century is useless because it is misleading. Something has to be done to close this dangerous knowledge gap.

General considerations

Given the diversity of collections in cultural institutions, digitisation raises many complex problems. However in order to encourage the digitisation of our cultural heritage, and build upon creativity for the benefit of future generations, all the varying issues raised by these differing kinds of works need to be addressed.

In order to comply with the norms that exist in intellectual property laws, digitisation requires the clearance of each and every copyright and related right. Therefore mechanisms have to be developed that recognise the rights and interests of the rights holders, but at the same time facilitate the digitisation that in turn will lead to greater innovation and creativity. Given the cost of rights clearance, it is in the public as well as private interest of European society to streamline rights clearance in a manner that is fair and balanced.

To this end, any solution must legally as well as morally and ethically address the following:

- The interest of the copyright holder who has invested time and money in creating copyright works and seeks to be remunerated for his or her investment.

- The interests of the rights holder who never created a work with commercial intent in mind, or has no further interest in commercialising his or her work, and simply wishes to share it with society.

- The moral rights of the creator who requires acknowledgement, and wishes his or her work to be used in line with the intent and purpose for which it was created.

- Cultural, political and religious sensitivities that are reflected in creative works produced by individuals or communities but lie outside of legally codified intellectual property laws.
The Comité did not try to cover the whole copyright debate, but decided to focus on the elements from its mandate: bringing orphan works and out of distribution works online. The recommendations presented here cover only these two issues.

For the orphan works, the Comité went beyond the terms of reference by also taking into account the need to avoid future orphan works and not just recommend a solution for existing orphan works.

**Orphan works**

There has been a lot of debate in the European Parliament and beyond about the best model to handle orphan works across the European Union. In addition, there has been a lot of debate in the Member States. Rather than promoting a specific model (exception to copyright, mutual recognition, European collective licensing), the Comité defined the conditions that need to be materially covered by a European instrument on orphan works and its implementation in the Member States.

The Comité believes that:

- **A European legal instrument needs to be adopted as soon as possible** to tackle the issue of orphan works. It welcomes the fact that such an instrument is in preparation by the Commission and has been announced in the Digital Agenda.

- In order to be effective, the legal instrument needs to fulfil all conditions of the following ‘8-step-test’. Under this ‘8-step-test’ the instrument should simultaneously:
  - **ensure that a solution for dealing with orphan works is in place in all the Member States.** Where no national instrument is in place, national legislation needs to be implemented.
  - cover all the different sectors: audiovisual, text, visual arts, sound.
  - **ensure cross-border recognition of orphan works:** An orphan work recognised as such in one Member State on the basis of a search in the country of origin, should be recognised as orphan across the EU. The Member State of origin is to be defined based on today’s geography, not on the basis of the historical borders at the time of publication. If there is no obvious country of origin, a search has to be conducted until a satisfactory solution is found.
  - **ensure the cross-border effect of this recognition:** an orphan work that is made accessible online in one Member State should also be made accessible online in all Member States or even globally.
  - **be compatible with the implementation of PPPs for digitisation.**
  - foresee, in the case of commercial use, a **remuneration for the rights holders** if after some time they are found or make themselves known. This remuneration can be kept in an escrow account. Also in the case of non-commercial use by cultural institutions the payment of an appropriate fee can be expected. For example, a one-off payment could be envisaged if the work is licensed under an extended collective license and the collecting society
has to indemnify the rights holders if they are found or make themselves known.

- ensure **reasonable transaction costs for dealing with orphan works**, commensurate with the commercial value of the work. This implies for example that a search for rights holders of older works can be less intense than for more recent works.

- be supported by rights information databases, such as the Arrow system currently under construction. These databases and a European list of established orphan works should be linked to Europeana as a key reference point for Europe's cultural heritage.

Preventing orphan works in the future is a main concern. In order to avoid orphan works in the future in an environment where creative production is exploding online (e.g. user generated content) without a clear indication of how to contact the creator, some form of registration should be considered as a precondition for a full exercise of rights.

The Comité realises that this would require a change in the Berne Convention and related instruments. Its members consider that a discussion on 'refreshing' the Convention should be taken up in the World Intellectual Property Organisation, and promoted by the European Commission.

Solutions based on registration also need to take into account privacy concerns of the creators, and need to be accompanied by awareness-raising amongst creators about the way in which they handle their rights.

Out of distribution works

The Comité believes that out of distribution works **need to be digitised** as part of Europe's cultural digital heritage.

The Comité favours solutions that provide the **widest possible online access** to out of distribution works for the users.

The solution must **cover all the different sectors**: audiovisual, text, visual arts, sound.

**Rights holders should be the first** to exploit out of distribution works.

The Comité distinguished three different cases:

- the material is **digitised for preservation purposes** by the cultural institutions and is available for consultation **on site**;

- the material is **digitised for preservation purposes** by the cultural institutions and made **widely accessible online**;

- the material is **digitised and commercially exploited by the rights holders**. In this case a digital copy should be deposited with the cultural institution that is responsible for preservation.

On funding the digitisation of out of distribution works the Comité distinguished between

- **investments in digitisation** for commercial exploitation of out of distribution works. These should be made primarily by the **rights holders**. There should be no direct public funding for digitisation in view of
commercial exploitation. Incentives like tax breaks for parties with commercial interest could be envisaged.

- investments in digitisation from the public sector. If the rights holders do not digitise the out of distribution works, digitisation should be paid for with public money. The digitised material should become freely available online, against an adequate remuneration for the rights holders.

Implementation issues:

- In many cases the rights for digitisation and online distribution will have reverted to or lie with the creators. Therefore the distributors must not automatically be seen as representing the rights holders. In some cases it can be the creators’ agents who must be addressed.

- in order to arrive at reasonable transaction costs for the digitisation of out of distribution works, it may be necessary to collectively manage the rights to older out of distribution works. This type of scheme should be backed by legislation. A date should be fixed to determine which works would fall under the collective management, which can be different from one sector to another.

- an adequate remuneration for the rights holders must be provided for and rights holders must always have the possibility to opt out from such collective schemes.

- in case rights holders do not exploit their material directly, collectively, or in collaboration with a private partner, cultural institutions should have a window of opportu-

- nity to digitise the material and make it available. This should be backed by legislation, e.g. those facilitating collective licensing solutions.

- whenever out of commercial distribution material is licensed to cultural institutions it should be freely accessible online. The rationale is that the public sector should not be seen as a competitor to the commercial distribution of in copyright works.

- whenever there is a choice between a solution where the material is licensed to libraries and freely accessible online, and a solution where the material is available against payment behind a pay-wall, the first solution is to be preferred.

- licences between cultural institutions and rights holders for bringing out of distribution works online should cover cross-border access, unless the extra cost for ensuring pan-EU access is disproportional.

- the solutions should be compatible with the implementation of PPPs for digitisation.

- solutions should be supported by rights information databases such as the ARROW system. These databases and a European list of licensed out of distribution works should also be linked to Europeana.

- solutions must also be foreseen for born digital material that is out of distribution.
Key recommendations

1. A European legal instrument for orphan works needs to be adopted as soon as possible and comply with the 8-step-test defined above.

2. Future orphan works must be avoided. Some form of registration should be considered as a precondition for a full exercise of rights. A discussion on adapting the Berne Convention on this point in order to make it fit for the digital age should be taken up in the context of WIPO and promoted by the European Commission.

3. National governments and the European Commission should promote solutions for the digitisation of and cross-border access to out of distribution works, taking into account the principles expressed above.

4. Rights holders should be the first to exploit out of distribution works.

5. For cultural institutions, collective licensing solutions and a window of opportunity should be backed by legislation to digitise and bring out of distribution works online, if rights holders and commercial providers do not do so.

6. Solutions for orphan works and out of distribution works must cover all the different sectors: audiovisual, text, visual arts, sound.
The issue

Europeana, Europe’s digital library, archive and museum, is probably the most ambitious cultural project ever undertaken at a European scale, bringing together cultural institutions from different sectors and from all the Member States. Europeana was launched as a prototype in 2008, giving access to some 2 million digitised objects, including books, maps, newspapers, journals, photographs, sound and video. In the span of less than two years, its collections have grown to more than 15 million digitised objects, thereby surpassing the initial target of 10 million digitised objects by the end of 2010. Currently, some 1500 cultural institutions contribute works to the site.

While Europeana receives strong support by the European institutions and individual Member States, it is fair to mention that contributions per Member State are still uneven. At the launch of the site, more than 50% of the objects came from French collections. Now France is still the largest contributor with 18% of the digitised objects, followed by Germany with 17% of the digitised objects. From some Member States there are only very few digitised objects accessible through the site and for many the key works defining their cultural and intellectual heritage are missing.

Europeana has had an important effect on discussions in the Member States on bringing cultural heritage online. The creation of the German Digital Library to be launched at the end of 2011 is for example a direct consequence of the development of Europeana. The site has also accelerated discussions between rights holders and cultural institutions in different countries on the best ways to bring in-copyright works online.

The Europeana site works as a portal and aggregator that gives access to the content stored de-centrally in the cultural institutions. Currently it does not host any content. The Europeana platform is based on open source software. In order to improve its services, Europeana is collaborating with a range of innovative European partners.

In the coming years, Europeana intends to develop its activities on the following priorities: 1) aggregating more content, 2) supporting the cultural heritage sector through knowledge transfer, innovation and advocacy, 3) distributing cultural heritage to the users wherever they are, whenever they want it, 4) helping users engage with their cultural heritage in new ways.

The knowledge about Europeana amongst European citizens, teachers and other potential users is still very limited.

General considerations

The Comité underlines the central place of Europeana in the strategy to bring Europe’s
cultural heritage online and to make cultural material available for work, education or leisure. It stresses the need to turn Europeana into the European cultural reference point online. This requires a concentration of financial efforts and political capital at European and at the national level for the development of the site. No parallel processes, projects or infrastructures should be funded at European level.

The Comité confirms the orientation of Europeana as a public project giving access to a wide range of trusted content from different types of cultural institutions. The site was not conceived as a competitor for digitisation projects in the private sector. Nevertheless, Europeana’s development should be evaluated against the background of the development of a private offer based on digitised material from cultural institutions across the world.

In a short time, Europeana has achieved excellent results. These results should be the basis for stepping up the ambition level for Europeana and transforming it into an endeavour on an industrial scale. In this context the Comité believes that Europeana will need to develop progressively and offer the services of an application platform rather than simply remaining a portal.

Furthermore, all main activities related to the digitisation and preservation of Europe’s cultural heritage should be linked to Europeana. An example is the ARROW database currently under development for orphan works which should be incorporated into the Europeana service. Europeana should also consider providing services to Member States that do not have all the tools in place.

**Enhancing the Europeana service**

The Comité commissioned an audit into the technical setup of Europeana, asking in particular whether the technologies underpinning the site and the technological development plan are adequate and future proof. The auditors concluded that the Europeana project is based on a robust technology foundation that includes open standards, architecture, and security. They suggested that Europeana looks into the possibilities of cloud computing and improves software measurement metrics, as well as the documentation of its software architecture.

The Comité feels that multilingualism is an essential area for development. Currently the Europeana interface is available in all the languages of the Member States. Cross-language search and automatic translation features should be incorporated in the medium term. It is also important to make translations of key works in different languages available. Furthermore there should be a clear link between works that go back to the same original, for example translations, adaptations or different editions of the same work.

European research and innovation projects in the area of research, multilingualism, digitisation technologies and social media should be encouraged to cluster around Europeana and contribute to the development of the site. In this way Europeana can become an important test bed for innovation and new ideas.

The Comité calls for a concerted effort by cultural institutions from all Member States to bring more digitised objects into Europeana. At the same time, the service should be en-
enriched with other content such as bibliographic data, tables of content and abstracts. In this case, it should be possible for the users to differentiate in the search results. They should for example have the possibility to get only results for digitised works that are fully accessible.

The Comité considers that it is equally important to make in-copyright content offered by private providers against payment searchable through Europeana. This offer would complement the material freely available. Partnerships between Europeana and private companies active in online distribution (e.g. publishers) should be encouraged. Also here, the users should have the possibility to filter the results for specific searches, if they prefer to find only the material that is fully and freely accessible.

**The Europeana infrastructure**

At the moment, Europeana does not store any material, but gives access to material stored de-centrally. It should be considered to give Europeana its own repository and archive, where it can host public domain digitised material (e.g. of institutions that have stopped their online activity). This has considerable cost implications, since storing information is expensive.

In the medium term, the Member States and the European Commission should consider turning Europeana into a deposit site for Europe’s digital heritage, keeping a digital copy of all digitised or born digital material produced in the European Union. For in-copyright works Europeana could be a dark archive functioning as a safe harbour.

In the short term, the repositories infrastructure funded through the EU Capacities Programme (e.g. the Driver project) could be mobilised in order to ensure that there are default repositories in each Member State feeding into Europeana.

**Recommendations to the Member States**

An efficient way to strengthen Europeana’s position as the reference point for providers of cultural content would be to channel funding for digitisation through the site. However, the Comité realises that Member States want to keep direct control over digitisation spending. The Comité is of the opinion that Member States should, as a minimum, ensure that all public funding for digitisation is conditional on the subsequent availability of the digitised material through Europeana.

Furthermore Member States should ensure that all public domain masterpieces of their cultural heritage are digitised and made available through Europeana within the coming 5 years. The Comité therefore strongly supports the ongoing development of a roadmap, undertaken jointly by the European Commission, the Member States and Europeana that will set a calendar for the Member States’ content contributions to Europeana and will help to monitor progress across Europe.

The Comité considers that the implementation of national aggregators feeding into Europeana is a positive step. All Member States should consider the creation of such aggregators.
Promoting Europeana

The Comité is of the opinion that a considerable effort is necessary to promote Europeana among the general public and in schools. To this end a specific part of the Europeana budget should be earmarked for the promotion of Europeana.

This is, however, not enough. The active promotion of Europeana is a responsibility for all, including cultural institutions, the European Commission and the Member States. In this context, the Comité notes that currently the websites of many cultural institutions which contribute content to Europeana do not link to the site. Such a link from the homepage of the website of cultural institutions is a minimum that can be expected.

The Comité notes that currently Europeana search results do not show up in searches in main search engines. Europeana should continue its talks with the search engines in order to rapidly overcome the barriers that are at the origin of this issue, since accessibility through the search engines will generate considerable supplementary traffic.

Key Recommendations

1. Europeana should be further developed to become the reference point for European cultural content online. This requires a concentration of financial efforts and political capital at European and at national level for the development of the Europeana site and the underlying structures.

2. Member States should ensure that all public funding for digitisation is conditional on the subsequent free accessibility of the digitised material through Europeana. They should also ensure that by 2016 they have brought all their public domain masterpieces into Europeana.

3. In the coming few years, Europeana should add to its portal an application platform, and main activities related to the digitisation and preservation of Europe’s cultural heritage should be linked to the site. In the technical development of the site particular attention should be paid to multilingual aspects. Europeana should also explore the opportunities of cloud computing in the future.

4. For the medium term, it should be considered to give Europeana a key role in the preservation of Europe’s heritage and to turn it into a European deposit site for public domain digitised cultural material and into a dark archive for in-copyright cultural material, both digitised and born digital.

5. Europeana must be actively and widely promoted by the cultural institutions, by the European Commission and by the Member States.
The issue

Digital preservation is a key challenge for the digital age. Digital materials have become an integral part of our cultural and scientific heritage. Due to rapid technical change, however, digital objects are more endangered than they appear. They are functional and meaningful only in specific hard- and software environments. When these environments change, the digital materials ‘age’ rapidly and are soon threatened by obsolescence and ultimately by loss.

The issue of digital preservation is very relevant in the context of digitisation projects. Today, only 22% of the cultural heritage institutions’ which digitise collections have long-term preservation plans in place. This means that the European investments in digitisation are in danger. Funding for digitisation that does not lead to sustainable digital assets can become a lost investment in the long term.

Digital preservation goes, however, far beyond digitisation and is also a core problem for any content produced in digital format. It therefore concerns practically all information produced today. Digital preservation issues cover the entire spectrum of digital objects, including raw data from research, text, websites, archival documents, art, film, music, databases and more.

The long term preservation of digitised and born digital material has four dimensions: organisational, legal, technical, and financial.

ORGANISATIONAL ISSUES

It is the planning for the long term, for the ‘indefinite future’, which makes high demands on organisation and planning. Preservation consists of a series of decisions that are made by different actors over time, e.g., the creator of a digital object, the provider, the curator. As short-term interests are often opposed to long-term goals, it is essential to clarify responsibilities at different stages throughout the lifecycle of a digital object.

Projects have delivered proofs of concept on national and on EU-level as to the technical framework conditions for digital preservation (e.g. PLANETS, CASPAR³). There is general consensus on the building blocks of long-term preservation systems, the principles of

---

¹ In this context ‘cultural heritage institutions’ means those cultural institutions that have a long term preservation task in their vision or legal obligations.
which are expressed in the OAIS standard\textsuperscript{4}. Lately, four large European national libraries have agreed on a description of the technical services needed for digital preservation.\textsuperscript{5}

Several Member States have digital preservation plans in place. Some Member States have established mechanisms to advise data creators on preservation aspects (Great Britain: Digital Curation Centre\textsuperscript{6}) and some Member States already have initiatives to coordinate opportunities for cross-sector cooperation and division of labour (UK: DPC\textsuperscript{7}, Netherlands, NCDD\textsuperscript{8}, Germany: nestor\textsuperscript{9}).

However, much work remains to be done in turning these strategies and principles into practice, even in the countries that are most advanced in this area. The countries that are lagging behind risk losing large quantities of digital content that will no longer be available for future generations.

**LEGAL ISSUES**

Legal deposit legislation has been or is being adapted to the digital revolution in a number of Member States. In many cases, however, gaps are left if not all sectors are covered. In Germany, there is, for example, a legal deposit obligation for online distributed digital text and music, including digitised material. Other sectors are not covered. In several countries the legislation does not yet foresee the possibility for web-harvesting, a process whereby the cultural institutions make a copy of the web in a certain domain in order to preserve the content.

Companies which operate across borders are often faced with the obligation to deposit the same content in several Member States. This content is then stored for preservation purposes by different cultural institutions across Europe. This raises questions about the cost, the administrative burden for the companies and the duplication of effort by the deposit organisations.

National copyright legislation does not always allow multiple copying for preservation purposes, which makes the work of deposit institutions in the digital age impossible. The digital preservation work is dependent on the migration of the content from one format to another in order to keep it alive and this is not possible if the number of copies the institutions can make is limited.

**TECHNICAL ISSUES**

Digital objects need persistent identification to guarantee their sustainable accessibility and avoid links leading to a void. Currently there is only a limited number of cultural institutions which use persistent identifiers. This poses risks for the long term preservation of the material and for its use by services such as Europeana.

Commercial content often functions only in combination with proprietary software, creating a major problem for archival institutions. Standardised and well documented file for-

\textsuperscript{5} http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/KB_Long_Term_Preservation_Services_2010-08-05.pdf
\textsuperscript{6} Digital Curation Centre: http://www.dcc.ac.uk/
\textsuperscript{7} DPC: Digital Preservation Coalition, http://www.dpconline.org
\textsuperscript{9} nestor: http://www.langzeitarchivierung.de/eng/index.htm
mats can be handled more easily than proprietary formats, but open documentation is not always in the interest of software companies.

The need to handle and preserve increasingly large volumes of dynamic data poses new technological challenges for institutions responsible for digital preservation.

**FINANCIAL ISSUES**

Digital preservation is expensive. The acquisition of digital material creates a long-term obligation on the host institution, which must be accounted for.

The emergence of cloud computing and other technical solutions can bring storage costs down in the coming years, but considerable financial and security challenges remain.

The Blue Ribbon Task Force on Sustainable Digital Preservation and Access has considered digital preservation from an economic perspective and come up with a set of recommendations, for example to strengthen incentives for digital preservation activities.10

**SUSTAINABILITY OF PORTALS**

When national or thematic cultural heritage portals are created, sustainable funding is not always ensured. A subsequent lack of funding for the portals can lead to their disappearance, which may have consequences for the visibility of cultural resources and may also have an effect on the development of Europeana that relies on thematic and sectoral aggregators.

The preservation of digital library environments such as portals is threatened by obsolescence, just as digital objects are. Adhering to open, widely distributed standards at the point of creation, documentation of design decisions, monitoring of technological change, and active enhancement and curation of the technology in use is important.

**General Considerations**

The private sector cannot be expected to take responsibility for the long term preservation of its content, once the commercial exploitation has ended. The Comité is therefore of the opinion that preserving our past and present cultural heritage is a public responsibility where cultural institutions have a central role to play.

Nevertheless, the companies which are driving the technological change must recognise their role in contributing to the obsolescence problem and work together with the cultural heritage institutions responsible for preservation.

Barriers preventing the institutions from carrying out their preservation tasks need to be tackled. There is no ‘final’ digital preservation solution, but a need to constantly monitor the technological environment and enhance preservation solutions to preserve digital data over the long term.

**Organisational issues**

In the view of the Comité, digital preservation requires additional and sustained effort on the part of policy makers, creators, pub-

---

lishers, producers, software companies and cultural institutions. Comprehensive preservation policies on European and Member State level are needed to protect the investment in Europe’s digital cultural heritage. All necessary framework conditions for digital preservation should be anchored in the respective national preservation policy.

National preservation strategies must address cross-sector cooperation and division of labour between possible custodians (e.g. National Libraries, National Archives, Data Preservation Centres) regarding the lifecycle management of all kinds of digital material, including audio and video.

Digital preservation strategies must be supported by the necessary organisational and technical infrastructures to be developed and implemented on the national level by all Member States.

In order to guarantee the preservation of the European digital cultural heritage, a system whereby a copy of all deposited cultural heritage material from the Member States is kept in a dark archive at Europeana should be considered.

**Legal issues**

The Comité stresses that copyright and other legislation has to enable the cultural heritage institutions responsible for preservation to create archival copies and to make file conversions for archival purposes.

The automatic process of harvesting needs a clear legal foundation on European level to enable efficient workflows.

To avoid duplication of effort by companies operating across borders and by the cultural institutions, a workflow could be envisaged by which any digital legal deposit copy that now needs to be deposited in several countries could be deposited only once. This system would include a workflow for passing on the copy to any institution that has a right to that copy under its respective national deposit legislation. Technical harmonisation measures on digital preservation at European level would facilitate the exchange of the content.

**Technical issues**

All digital objects in European digital libraries should be persistently identifiable. Member States should encourage cultural institutions to use persistent identifiers in digitisation projects. This will contribute to preserving the digitised content and to improving the search (by overcoming the problem of broken links in Europeana and by creating the possibility to establish connections between the objects). Reliable resolution services like the Europeana Resolution Discovery Service are of the utmost importance in this context. A reliable resolution service for persistent identification of digital objects must be developed and maintained on European level, preferably linked to Europeana.

Reliable technical information about the digital objects helps in planning preservation actions. The development of file format registries supports this process and must be intensified to support preservation planning.

All digital preservation solutions should be transparent and scalable in order to adapt to
the continuous technological development and the growing amount of data.

Mechanisms have to be developed to solve the conflict of interest in open documentation between the software industry and the cultural heritage institutions responsible for preservation.

Research & Development initiatives with a strong reference to practical user needs have to be supported on national and European level to constantly monitor the technological environment and to enhance preservation solutions. In addition, further research into solutions for handling large volumes of dynamic data is necessary.

Financial issues

During the planning phase of digitisation projects the cost of preserving the digital assets must already be duly taken into account.

Digital preservation requires continuous effort and adequate financial resources are needed at national and European level. In addition to funding targeted research projects, sustainable digital preservation budgets have to be in place. Although there is a high recognition of the importance of digital preservation research in several Member States and on the European level, the need for financing the day-to-day process of digital preservation is not yet appropriately reflected in institutional budgets.

The area of digital preservation should not only be considered from the point of view of cost. It also has an enormous economic potential. The efforts and experience of the cultural institutions in improving the efficiency of preservation can be an important asset in the fast growing market for storing digital data.

Sustainability of cultural portals

The sustainability of digital library portals, especially the national and/or sectoral aggregators feeding into Europeana must be ensured.
Key Recommendations

1. Preservation is a key aspect in digitisation efforts. Digital preservation is also a core problem for any born digital content. The organisational, legal, technical, and financial dimensions of long term preservation of digitised and born digital material should be given due attention.

2. To guarantee the preservation of the European digital cultural heritage, a copy of digitised or born digital cultural material should be archived at Europeana. For in-copyright works the deposit site would be a dark archive functioning as a safe harbour.

3. To avoid duplication of effort by companies operating across borders and by the cultural institutions a system could be envisaged by which any material that now needs to be deposited in several countries would be deposited only once and then passed on to every cultural institution that has a right to it under its respective national deposit law.

4. Copyright and related legislation has to enable the cultural heritage institutions responsible for preservation to create archival copies and to make file conversions for archival purposes.

5. Persistent identifiers must be implemented in each digital object archived in cultural institutions. A reliable resolution service for persistent identification of digital objects must be developed and maintained on European level, preferably linked to Europeana.
The issue

THE COST OF DIGITISATION

The terms of reference of the Comité underlined the need to assess the overall financial cost of digitising Europe’s cultural heritage. To this end, the Comité engaged the help of a specialised contractor. The report of the contractor is presented as an Annex. It builds on earlier extensive work done in this area at European and at national level.

The report concluded that the estimated cost of digitising the total collections of Europe’s museums, archives and libraries, including the audiovisual material they hold, is approximately €100 billion. The Comité does not have the ability, competencies or means to assess the accuracy of the numbers.

This type of calculation is necessarily based on a series of hypotheses and extrapolations. Nevertheless it indicates a plausible order of magnitude of the funding needed to digitise the whole of our cultural heritage, commensurate with the richness of European culture and civilisation over the last centuries.

The €100 billion can be divided as follows over the collections of the different types of cultural institutions.

<table>
<thead>
<tr>
<th></th>
<th>Total estimated mean cost of digitisation (in Bn €)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library collections</td>
<td>19.77</td>
</tr>
<tr>
<td>Museum collections</td>
<td>38.73</td>
</tr>
<tr>
<td>National Archives</td>
<td>41.87</td>
</tr>
<tr>
<td>Audiovisual archives</td>
<td>4.94</td>
</tr>
<tr>
<td><strong>Gross total</strong></td>
<td><strong>105.31</strong></td>
</tr>
<tr>
<td>Cumulative efficiency gain (if digitisation takes place over 10 years)</td>
<td>-5.3</td>
</tr>
<tr>
<td><strong>Net total</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

The figure includes the digitisation of 77 million books, 24 million hours of audiovisual programmes, 358 million photographs, 75.43 million works of art, 10.45 billion pages of archives. The collections that have already been digitised or that have been excluded from digitisation for a specific reason are not integrated in this number. According to the report this is what needs to be digitised.

---

1 These are explained in the methodological part of the report on the cost of digitisation.
The report notes that, although comparisons should be made with caution, investments of this type are common in other policy areas. For instance, the Research & Development Budget for the Joint Strike Fighter programme is estimated at €40.34bn, which represents 40% of the amount needed to digitise the whole of Europe’s cultural heritage. Equally, the funding needed for building 100 km of main road is equivalent to the cost of digitising every piece of audio content in EU cultural institutions, or of 16% of all books (individual titles) in EU libraries.

The report also underlines that mass-digitisation is an industrial process, and hence very susceptible to efficiency gains due to scale. Furthermore 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.

It is difficult to give an accurate overall estimate of the current investments in digitisation across the European Union. All Member States have reported to the European Commission on their plans for and investments in digitisation.2 However, the figures concern mostly the digitisation effort at national level, without necessarily taking into account the funding by the individual institutions or at local and regional level, or the private funds invested in digitisation in the context of public-private partnerships.

As underlined in the preceding chapter, the acquisition of digital material creates a long-term obligation on the host institution. Preserving information and keeping it accessible is an expensive activity.

FUNDING EUROPEANA
In its start-up phase, the Europeana service has been largely driven by project-financing from the Community programmes covering some 80% of the costs. The ‘Europeana 1.0’ project was co-funded with €6.2 million from the eContentplus programme and will run until mid-2011. Several Member States are providing co-funding: the Netherlands, Germany, France and Spain are the main contributors, with smaller contributions coming from a range of other Member States, as well as Norway and Switzerland.

In the period until 2013 the Commission will continue co-funding Europeana through projects (some 80% of the costs). An amount of €9 million has been allocated through the CIP programme for a major project that will support the further development of Europeana in the period mid-2011-end 2013. For this period further co-funding by Member States or from other sources will be necessary. Several Member States have already indicated that they will contribute.

Funding an endeavour such as Europeana through a myriad of projects, complemented by voluntary funding by the Member States presents serious problems in terms of continuity of development. Therefore the Commission has announced in the Digital Agenda for Europe that, by 2012, it will make a proposal for a sustainable funding model for Europeana. The Resolution of the European Parliament and the Council Conclu-

2 These reports can be found at the following web-address: http://ec.europa.eu/information_society/activities/digital_libraries/other_expert_groups/mseg/reports/2010/index_en.htm
sions on the next steps for Europeana - both texts were adopted in May 2010 - reaffirmed the need for a thorough reflection and a clear proposal on funding Europeana after 2013.

**ECONOMIC BENEFITS OF DIGITISATION**
Digitisation represents a considerable financial investment. At the same time it will be at the basis of large economic opportunities. There are three main areas where digitisation can help nurture and stimulate economic growth and job-creation.

The first is, of course, the digitisation process itself and the technologies directly linked to it. Part of the digitisation tasks will probably be outsourced to private companies. And these tasks will require equipment such as advanced scanners (e.g. for 3D rendering), as well as enhanced software, for example for optical character recognition. If European companies can develop the most efficient technologies and working methods, they will be the first to benefit from public contracts for digitisation. The process of digitisation, however innovative or advanced the tools, will be labour intensive and would result in job creation in large numbers.

Second, the value of the digitised material should not be underestimated. Digitised cultural content can become an important raw material for services and products in areas such as tourism, education and new technologies. In an environment where the market for mobile applications is growing rapidly - some estimations expect the global market for apps to grow to $32 billion in 2015 ³ - widely available cultural material will be a key asset for new services. Although not all the potential usages of the digitised material can be predicted today, we can reasonably expect that the arrival of millions of new digitised cultural works online will spur a wave of innovation and new business models for companies specialised in various stages of the digitisation chain, as well as in creative content.

An example of such emerging business models was presented at the public hearing on 28 October 2010. Arkhopôle, based in the French Aquitaine region, is a cluster of 125 SMEs specialised in the creation and commercialisation of cultural content, collaborating with cultural institutions and universities. Its vision is to develop in France and Europe a new market for digital heritage material.

The economic value of the digitised material lies not only in the possibility to directly integrate it into new services, but also in the associated new knowledge and information. An example is the new insights necessary for developing language technologies gained by Google as a result of the Google Books project. Other players may join this and related areas tomorrow and create value through the digitised content.

The third area that represents a considerable economic potential is related to the storage, preservation, and processing of digital material. These activities face a new step in their evolution, as cloud computing is gaining speed and companies are managing bigger and bigger databases. One indication is a recent bidding war in this area.⁴ It is likely that

---

³ Juniper Research.

⁴ In a recent bidding war between Dell and HP over 3Par, a data storage company, the firm with only 650 employees was bought by HP for $2.35 billion. The Economist, 6 November 2010, 'A special report on smart systems', p. 12.
the importance of this sector will grow tremendously in the coming years and the handling of digitised and born digital cultural material by heritage institutions and private companies is an important field for experimentation and innovation.

General considerations

Before setting out its position on the funding of digitisation and Europeana, the Comité would like to reiterate that all the material available for digitisation should be digitised as long as it can be considered part of our cultural heritage.

In addition, the Comité would like to underline again that it favours the widest possible access to the digitised items.

The financial challenges of bringing our cultural heritage online are huge, but so are the potential rewards. These rewards should not only be defined in terms of better access to cultural content, but also in economic terms, even though they cannot be accurately measured or forecasted.

Funding digitisation

A first question relates to the respective responsibilities of the public and private sectors for digitising our cultural heritage. Considering that access is the main ground for digitisation, the position of the Comité is that the public sector has the primary responsibility for making our cultural heritage accessible and preserving it for future generations. It should therefore bear the largest part of the costs involved. The control over Europe's heritage cannot be left to one or a few market players.

However, the task of digitising Europe's cultural heritage is gigantic and therefore the Comité thinks that the private sector must be involved in funding the digitisation effort. In fact, a wide range of private players could potentially be interested in (co-)funding digitisation: media companies, the digitisation industries (a rapidly growing market), internet, telecoms and IT companies interested in new digital content and services, the tourism and learning sectors, not forgetting other companies which would like to be associated with culture for reasons linked to their image. Conditions for such public-private partnerships for digitisation are treated in the next chapter.

Another question is whether the public digitisation effort should take place at national and regional level or whether there should also be a major investment at the European level. The Comité believes that the digitisation of collections in the Member States is primarily the responsibility of national or regional governments. Hence funding should be made available from budgets within Member States. However, Member States should be strongly encouraged to use the funding possibilities from the European Structural Funds for digitisation activities. Also, some targeted digitisation efforts with a clear cross-border scope (e.g. cross-border or multilingual collections) could be co-funded at European level.

Neither the size of the task of digitising our common heritage, nor the current financial crisis can be an excuse for not acting. The Comité therefore calls on all Member States to increase their budgets for digitisation and to reduce the gap between the investments needed to digitise cultural collections and the current level of funding.
The available budgets must be used in the most efficient way. This requires in the first place avoiding that the same object is digitised several times. Moreover, instead of leaving each cultural institution alone with its obligations, digitisation should be organised and strategically planned at national level in order to maximise efficiency gains and facilitate synergies and the exchange of good practices. Each Member State should be required to elaborate a strategic plan and these should be shared at the European level.

The Comité decided not to impose selection and not to encourage choosing among the different categories which type of works is more entitled to digitisation than another for the sake of cost reduction. Nevertheless it is obvious that priorities will have to be defined and the task will have to be organised and scheduled over several years, or even decades. Information on the priorities should be exchanged at the European level in the existing Member States’ Expert Group on Digitisation and Digital Preservation, in order to achieve synergies.

The cost of preserving the digitised material must be accounted for and addressed from the outset in programme budgets.

Funding Europeana

The cost of Europeana is extremely low compared to the budgets needed for digitisation and taking into account its potential to make the digitised material representing our common heritage more visible and accessible to all.

Given Europeana’s character as a common good, the Comité is of the opinion that public funding should cover the largest part of Europeana’s operational costs, also after 2013. To this end, the funding of Europeana should be considered together with the investments of Member States in digitisation and in setting up national aggregators. In other words: The funding of digitisation and of Europeana should be seen as a package, where Member States are broadly responsible for funding the digitisation of their cultural heritage and where the funding of the Europeana portal should come predominantly from the budget of the European Union in view of its European added value.

Because of the financial crisis, this is not the best moment to attract private sponsors or investments for Europeana. Companies would expect a business return based on hard visitor numbers and profiles and/or a clear link between the site and the brand. However, for the medium term, complementing the public funding for Europeana with some form of sponsorship or partnerships with business seems to be a realistic perspective, e.g. sponsorship by Telco or IT companies operating on a Europe-wide basis, or companies who want to be associated with culture.

A clear vision and plan - going beyond a mere business plan - is necessary for the further development of Europeana. The vision and plan should include the objectives to be achieved for the medium term and indicate the steps that will be taken to turn the site into an indispensable reference point for European culture online.

Reaping the economic benefits of digitisation

Digitisation activities must be undertaken with an open eye for the business opportuni-
ties involved. It is the conviction of the Comité that the organised combination of public funding and private investment could lead to unlocking new growth opportunities that make digitisation an economic opportunity.

Implementing innovative economic models for digitised cultural material is an area deserving special attention. For instance, local and regional clusters of SMEs specialised in the creation and exploitation of digital heritage, in partnership with cultural institutions, could help create a niche for local growth and development, ranging from digitisation of cultural assets to the re-use and exploitation of the digitised assets. SMEs can contribute specific competences in the digitisation chain.

Fostering ‘knowledge’ partnerships between cultural institutions and universities in order to boost research and innovation in digitisation of cultural assets is another area that should be further explored.

Developing strategic partnerships at European and international level in the area of new technologies and applications in relation to cultural heritage can help transfer knowledge and build capacities at local, regional and national level.

One of the areas that have a great economic potential is digital preservation. The emergence of strong European players in this rapidly growing area should be encouraged.

Bringing out the full value of the digitised assets and the new opportunities linked to digitisation and preservation activities presupposes that the actors in the private and public sectors have the necessary skills. Identifying the training needs of professionals involved in the digital economy and creative content online and providing specialised professional training can help to develop competences matching current and future needs.
Key Recommendations

1. The public sector has the primary responsibility to fund digitisation, and Member States will need to considerably step up their investments in digitisation. The current financial crisis cannot be ignored, but equally cannot be a reason for not acting.

2. The involvement of private partners should be encouraged. Private funding for digitisation is a complement to the necessary public investments and should not be seen as a substitute for public funding.

3. Digitisation should in principle be funded at the national or regional level, not at the European level. However, the Member States should be strongly encouraged to use the funding possibilities from the European Structural Funds for digitisation activities. Also, some targeted digitisation efforts with a clear cross-border scope (e.g. cross-border collections) could be co-funded at European level.

4. Given Europeana’s character as a common good, public funding should cover the largest part of Europeana’s operational costs, also after 2013. The funding of digitisation and of Europeana should be seen as a package, where Member States are broadly responsible for funding the digitisation of their cultural heritage and creating national aggregators and where the funding of the Europeana portal should come predominantly from the budget of the European Union.

5. Member States should promote ways to turn digitisation into new development opportunities for European firms, for example through regional clusters of businesses in partnership with cultural institutions, knowledge partnerships between cultural institutions and universities, or through strategic partnerships at European or international level in the area of new technologies and applications in relation to cultural heritage.
Public-private partnerships for digitisation opportunities, challenges and pitfalls

The issue

Different models of public-private partnerships for digitisation of cultural assets exist in Europe. They range from direct investment of funds in return for exclusive commercial exploitation of the digitised material to classical sponsorship schemes for advertising/marketing purposes.

Thus far, private funds have been largely directed at the digitisation of books and concern public domain material. A closer look at the situation on the ground reveals some subtle, yet quite persistent, messages as regards the conditions of public-private partnerships for digitisation in Europe.

Google has emerged as a major player globally through its massive investment in digitisation of books, but also in Europe through a number of partnerships with European libraries. In general, Google estimates that, out of 130 million unique books existing in the world, it has digitised 15 million since the launch of the ‘Google book’ project in 2004. Google digitises only public domain material in Europe (earlier than 1870). There has been no disclosure as regards the exact funds invested or the nature of agreements with libraries. Nonetheless, on the basis of public statements linked with two recent agreements some indications arise:

- Under the agreement with the Austrian National Library, 400,000 books will be digitised by Google. An investment of €30 million (around €75 per book) was reported by the Austrian side.

- Under the agreement with the Italian government, 1 million books will be digitised. The Italian side estimated the investment by Google at around €100 million (€100 per book), while Google has argued that the invested amount would be ‘much, much less’ and ‘six to ten times lower than what other potential partners have envisaged’.

Access to digitised material is free for the end-users and all material may be used for non-commercial purposes. Libraries receive copies of the same quality. The period of preferential use of the digitised material is discussed on a case by case basis de-
pending on the size and type of collection. Some agreements (i.e. Lyon library) foresee 25 years, but this has progressively come down to 15 years.

Although Google has not been explicit about its business model, enhancing the quality of its services as a search engine seems to be a core business objective. The recent launch of ‘Google e-books’ - based on agreements with publishers for bringing in-copyright works online - casts a new light onto Google’s ambitions in the digital book market.

The British/American publishing group ProQuest has been active in the digitisation of public domain books as part of their project Early European books until 1700. Agreements have been signed with libraries in the UK and the National Library of Denmark (2009). According to the agreed business model with the latter, ProQuest maintains the database for sale (target audience: universities and researchers) and the library receives a copy for immediate dissemination within Denmark (controlled by IP addresses). Ten years after its publication by ProQuest, the digitised material can be made available to everybody.

Another type of public-private partnership concerns agreements between cultural institutions and private partners who are not interested in exploiting the digitised content, but in enhancing their corporate image. For example, sponsoring by Telefónica for the digitisation of the collections of the Spanish national library was reported to amount to €10 million (for the digitisation of 25 million pages). It is worth noting that tax incentives for this kind of partnership exist in Spain, as well as in other EU Member States.

General considerations

Stimulating the flow of private funds for the digitisation of cultural assets through equitable public-private partnerships appears as a viable and sustainable way of tackling the pressing question of making Europe’s cultural wealth accessible online and preserving it for future generations. This affirmation becomes all the more acute in the wake of the current financial downturn and the growing pressure on public budgets.

The key question is not whether public-private partnerships for digitisation should be encouraged, but ‘how’ and ‘under which conditions’. Which conditions and principles should apply in order to reconcile the legitimate interests of the parties involved, namely the cultural institutions, the commercial parties, and the right holders in case of digitisation of in-copyright material? How to promote a level playing field and stimulate involvement of European actors? How to help build strategic partnerships for digitisation at local, regional and cross-regional level?

Which conditions and principles for public-private partnership for digitisation in Europe?

Access to Europe’s cultural resources for all is also relevant in the case of public-private partnerships and this principle should not be compromised by the necessity of promoting
Europe’s global competitive edge in the digital era.

As guardians of cultural heritage, aggregators of cultural content and educators, cultural institutions are obliged to safeguard their core mission, whilst reshaping their traditional role against the backdrop of the digital shift. Making their collections accessible online enables them to seize the opportunities of the digital era. Nonetheless, entering into a partnership with a private entity should not result in compromising the core mission of cultural institutions, but rather further allow them to reap the full benefits from the digitisation of their collections in terms of online accessibility, preservation and re-use.

On the other hand, private partners are a crucial part of the digitisation chain. They bring funding, technology and expertise. It is logical that they wish to see returns on their investment.

Striking the right balance in public-private partnership agreements is a question of fine negotiation between cultural institutions and private partners and varies from one case to another. Cultural institutions often enter partnerships with commercial entities ‘unprepared’ or ‘unequipped’.

In this environment, it is useful to set out basic conditions and principles that could apply when private funds are invested to digitise and make accessible online collections of European public cultural institutions:

- **respect for right holders**: public-private partnerships should be established within the applicable EU and national copyright legal framework.

- **transparency**: both the process for reaching the agreement and the contents of the agreement between a public cultural institution and a private partner should be made public. This is a minimum requirement for the digitisation of public cultural assets.

- **access**: partnership models where the end-user has **free access** to the digitised material should be clearly encouraged over models where the end-user has to pay for accessing the material. This should apply both to public domain and in-copyright material. When entering into public-private partnerships for digitisation, the cultural institutions should ensure **cross-border access** to the material. Partnership models giving citizens from one EU Member State more favourable access conditions to the digitised material than citizens from another Member State should be strongly discouraged.

- **quality of copies**: the private partner should provide cultural institutions with digitised files of the same quality as the ones it uses itself.

- **re-use**: the cultural institution should be able to use the files without restrictions in non-commercial contexts.

- **revenue sharing schemes**: in the context of public-private partnerships, cultural institutions could negotiate revenue sharing schemes in relation to the commercial exploitation of the digitised material (ie. advertising).

- **exclusivity**: in principle there should be **no exclusive partnerships**. In other words, public domain material that is the subject
of a public-private partnership may be also be digitised by other private companies.

**Preferential use or preferential commercial exploitation:** it is logical that the private partner seeks a period of preferential use or commercial exploitation of the digitised assets in order to avoid free-rider behaviour of competitors. This period should allow the private partner to recoup its investment, but at the same time be limited in time in order to avoid creating a one-market player situation. For these reasons, the Comité set the maximum time of preferential use of material digitised in public-private partnerships at **maximum 7 years**. The seven-year time span is considered adequate to generate, on one hand, incentives for private funds’ investment in mass digitisation of cultural assets and, on the other, allow sufficient control of the public institutions over their digitised material.

How to stimulate a flow of private funds and create a level playing field in the digitisation of Europe’s cultural heritage?

Despite the varied landscape of options and possibilities, in some areas, such as the digitisation of public domain books, some players have a more dynamic presence than others. Global players have a lot to offer, but they should not be seen as a one-way solution to the digitisation of Europe’s cultural assets. A search for a European solution means seeking innovative ways of attracting private funds, building expertise and creating the conditions for growth.

Policy makers should consider enabling the involvement of European players through different means and measures:

- Encouraging digitisation in areas that have not received much attention thus far, such as audiovisual material, newspapers, periodicals or museum objects (as opposed to books) may help diversify the landscape of digitisation in Europe.

- In the medium term, subject to an improvement of the financial situation in the Member States, creating incentives through taxation is to be considered.

- The use of matching public funds for digitisation is another avenue to explore. Public funds may be given to cultural institutions which have secured a partnership for the digitisation of their collection with a private entity, on a matching basis with the private funds invested, thereby strengthening the scope and output of the digitisation project.

- Encourage Europeana and its contributing institutions to expand their digital contents by building partnerships with European businesses.
Key Recommendations

1. In order to protect the interests of public institutions entering into a partnership with a private partner a set of minimum conditions should be respected:
   - The contents of the agreement between a public cultural institution and a private partner should be made public.
   - The digitised public domain material should be free of charge for the general public and available in all EU Member States.
   - The private partner should provide cultural institutions with digitised files of the same quality as the ones it uses itself.

2. The maximum time for preferential use or preferential commercial exploitation of material digitised in public-private partnerships must not exceed 7 years. This period is considered adequate to generate, on one hand, incentives for private investment in mass digitisation of cultural assets and, on the other, to allow sufficient control of the public institutions over their digitised material.

3. Policy makers at European and national level should create favourable conditions for the involvement of European players. In particular:
   - Encourage digitisation in new areas that have not received much attention thus far, such as audiovisual material, newspapers, periodicals or museum objects.
   - In the medium term, subject to an improvement of the financial situation in the Member States, create incentives for the investment of private funds through taxation.
   - Encourage the use of public funds matching private funds invested in digitisation. Public funds may be given to cultural institutions which have secured a partnership for the digitisation of their collection with a private entity, on a matching basis with the private funds invested.
   - Encourage Europeana and its contributing institutions to expand their digital contents by building partnerships with European businesses.
Having reached the conclusion of the present report, we hope that we have been able to illustrate how important it is for Europe to collectively ensure that our cultural and intellectual heritage is accessible to all by means of new information technology.

A danger looms on the horizon: that we lose parts of our heritage for the simple reason that it cannot be consulted with the methods and tools used today and which we can expect to further expand in future. This danger is particularly preoccupying for Europe. For the role of Europe in the world, whilst endangered by economic, let alone political regression, will remain a motor in the cultural domain - provided that culture keeps pace with current and future developments.

A poet said: ‘Europe, thy name is memory.’ Europe was constructed with the notion of evolution, thought, creation, research and ingenuity. No one will disagree: each phase of this process is worthy of conservation and study. It presupposes today a gigantic translation of this collective heritage to ensure that it will not become as impenetrable for future generations as have been the hieroglyphs throughout the centuries! It is above all the duty of the public sector to ensure that this heritage is available to everyone.

A number of initiatives have been taken, of course. They are the results of sporadic, dispersed measures adopted with a piece meal approach. But there is one initiative that has emerged, resulting from the converging efforts of different countries and a large number of institutions: Europeana. It has proved to be a useful, timely and adequate initiative, but has so far remained modest in its development. Wherever Europeana takes root, the network provides excellent services. These efforts however are unfortunately still too fragmented, and overly dependent on the goodwill of its stakeholders. Europeana should have greater means, coming from the European institutions and the Member States, taking account of the principles of subsidiarity and representativity.

There is another reason to support the efforts made by Europeana so far: the respect of some fundamental principles, of which Europe is the guarantor in the area of culture. They are principles that represent its contribution to civilisation: first, the specific status of cultural activities which should not be entirely subject to market rules; and, second, the protection of the rights of the creator, too often neglected or even disdained nowadays, and left to a large extent unprotected from abusive practices of unauthorised exploitation.

So what would happen if we did nothing to preserve our heritage in the digital era? What if we decided to leave the cultural works as they are without transforming them into digital formats and without entering into any concerns of preservation and sustainability? Would our inaction have a ‘cost’?
First, protecting and promoting culture and heritage as a public value is a shared responsibility and one founded in the EU Treaty: this means championing our cultural diversity and bringing our common heritage to the fore; bridging tradition and modernity; reconciling youth with the past, thereby paving the way for the future; providing a solid basis for education and knowledge dissemination; enhancing the social fabric and strengthening social cohesion; contributing to knowledge dissemination in the information society.

Moving beyond the moral register, culture and heritage in the digital era represent a set of opportunities for European economies and societies. Digitisation relies on technological progress, but, in turn, may also spur innovation and creativity. It can contribute to job creation, growth and business development in sectors linked to technology, culture, creativity and innovation. Addressing the economic dimension of the digitisation of our cultural heritage does not undermine the symbolic dimension, but rather helps us make the most of the enormous wealth we have in our hands, which is too often left untapped. Our heritage and culture can reinforce Europe's competitive edge in the era of the digital revolution and globalisation. Innovative business models, smart investments, collaboration between sectors (ie. public-private, cultural-business, creative-technological), policies adapted to the needs of stakeholders (ie. cultural institutions, creators, private partners, the general public) can help tackle the transition to the digital era in a dynamic and forward-looking way.

Time is of the essence. Global competition is harsh and Europe risks missing out on the infinite opportunities offered by the digital shift.

A final thought: our reflection focused on the conservation and valorisation of our heritage, but it is not looking solely at the past. By honouring our heritage, we prepare fertile ground for future growth.
Annexes

Annex One: Terms of reference of the Comité des Sages on bringing Europe’s cultural heritage online ....................... 55

Annex Two: Overview of the results of the online consultation ...... 61

Annex Three: Summary Record of the Public Hearing of the Comité des Sages on bringing Europe’s cultural heritage online ........................................................................................................ 85

Annex Four: Results of the Cap Gemini technical audit of Europeana ................................................................. 105

Annex Five: Report of the study into the cost of digitisation by the UK Collections Trust .............................................. 111
Terms of reference of the Comité des Sages on bringing Europe’s cultural heritage online
Terms of reference of the Comité des Sages on bringing Europe’s cultural heritage online

Introduction

The European Union faces the need to address digitisation of Europe’s cultural heritage and cultural content in a comprehensive and diligent manner. Following the discussions by EU Ministers at the Education, Youth and Culture Council of 27 November 2009, the Commission has decided to set up a Reflection Group to develop a vision and address the main issues at stake in bringing Europe’s cultural heritage online.

A large part of Europe’s rich and diverse cultural heritage is held by libraries, archives and museums across the continent. The digitisation and online availability of cultural material - including books, maps, newspapers, sound, archival material, photographs, museum objects and audiovisual material - will make it more accessible for Europe’s citizens and easier to use for leisure, work and study. The material can also be an important input for new creative efforts and for a wide range of information products and services, for example in sectors such as education and tourism.

At the same time, Europe’s artists and creators are working in an environment increasingly shaped by digital technologies and global communications. The digital shift shakes up traditional models, transforms value chains and calls for new business models. The digital transition raises a number of significant challenges for the creative sectors. They indeed need to adapt to new opportunities brought by digitisation in terms of creation, production, diffusion and exchanges of new cultural expressions to fully benefit from them.

It is therefore crucial to reconcile the traditional mandate of cultural institutions, which give access to our common heritage and preserve it for future generations - with the transition towards the digital age. Managing the ‘digital shift’ requires a strategic approach that will ultimately define how European citizens will access and use digitised cultural resources, and eventually how European creativity will be preserved and promoted for future generations.

The challenges of putting Europe’s cultural heritage online

In 2005, the European Commission launched the digital libraries initiative, with the aim to make Europe’s cultural heritage accessible to all on the Internet. One of the tangible results of this initiative is the creation of Europeana...
In order to make this project a success, a number of challenges have to be addressed. One of the key issues is who pays for digitisation and the implications this has for control over and responsibility for our digitised cultural heritage, including the respective roles of the public and private sector. The choice of funding models will have a fundamental impact on the access to the digitised material and the way it can be re-used in the information society.

Public funding for digitisation ensures that libraries, archives and museums keep the control over the digitised material, and that they can use it to fulfil their remit of making information accessible and preserving it for the future. However, digitisation is expensive and the collections to be digitised are vast. Therefore cultural institutions are often required to recover part of the digitisation costs. They often do so by charging for the commercial re-use of the digitised material, sometimes by charging online access by the public.

When the digitised assets are in the public domain - *i.e.* no longer covered by copyright - this poses a fundamental dilemma. Recuperating (part of) the costs may be necessary for the cultural institutions to engage in further digitisation, but at the same time charging for access and re-use locks up public domain material and limits its usefulness in the information society. Such models run the risk to contradict the basic aim for which the material was digitised in the first place. Alternative models, for example those that involve advertisements on the websites of the cultural institutions to generate some income, could endanger the neutrality of the cultural institutions as information providers.

The situation is equally complicated for in-copyright material, which covers most of the collections of the 20th century. A large part of this material is no longer commercially available, and for many works it is even impossible to locate the rights holders (the so-called ‘orphan works’). For the digitisation and online accessibility of this material, there are two basic issues. The first one concerns the roles of respectively the cultural institutions and the private sector (both rights holders and technology firms) in digitising this in-copyright material. The second one concerns the transaction costs for digitising older works, which are at present prohibitively high. There is a risk that the uncertainty about what a ‘fair price’ is and/or that the high transaction costs for rights clearance will lead to inertia from the side of the public institutions and or rights’ holders. Without more simple rights clearance mechanisms, large scale digitisation of out-of-print works and orphan works is therefore unlikely to happen, which may result in what has been described as a ‘20th century black’ hole in European collections online and more in general on the Internet.

Another source of funding for digitisation is the private sector. Public-private partnerships, complementing public investments in digitisation are a promising way forward to bring Europe’s cultural heritage online. They raise, however, a range of new challenges. Public-private partnerships often entail a cer-
tain control over the digitised material by the private partner who expects a return on investment through the exploitation of the content. Without this incentive private partners may refrain from investments in digitisation. As a result, material that is no longer covered by copyright may be de facto removed from the public domain by limiting access for the end-user and/or by limiting the use of the digital material by potential competitors. The partnerships raise further questions, for example on exclusive rights granted by public sector bodies to private partners, and on responsibilities for preserving the material over time, which may entail costs that are higher than the actual digitisation costs. Cultural institutions across Europe are struggling with these questions when they are approached by private companies who offer to digitise their collections.

Finally, digitisation - regardless the source of funding - gives rise to important questions on the conditions set to access the material once it has been digitised, and on the format used to ensure its long-term preservation. An overview of best practices in Europe and elsewhere would help avoid the development of fragmented approaches: national silos of online cultural resources could indeed lead to discriminatory conditions of accessibility for European citizens and difficulties in interoperability of the IT systems in place. That could in turn jeopardise the vision and services of Europeana, as the common access point to Europe's cultural heritage.

Terms of Reference

The Group will provide a set of recommendations for the digitisation, online accessibility and preservation of Europe’s cultural heritage in the digital age, looking in particular at the issue of public-private partnerships for digitisation in Europe. The recommendations will need to consider the variety of ongoing policy initiatives at the EU1 or Member States’ level, including the large scale funding of digitisation by public authorities and, where relevant, legislative initiatives.

It should draw upon analytical work done at the national level and identify best practice examples in the Member States or in Third Countries, to highlight possible solutions for questions of principle as much as for more practical problems of technical compatibility.

The analysis of the Group should build on the work of the High Level Group on Digital Libraries that was active between 2006 and 2009, and take into account ongoing discussions at European level, such as the work on public-private partnerships carried out in the context of the Conference of European National Librarians.

The work of the Group will cover in particular the following areas:

- the overall financial cost and level of public funding available to European cultural institutions for digitising their collections in Europe;

---

1 The Commission has addressed related issues in a Recommendation on 24 August 2006 on the ‘digitisation and online accessibility of cultural material and digital preservation’, and more recently in the Communications of 28 August 2009 on ‘Europeana next steps’ and of 19 October 2009 on ‘Copyright in the knowledge economy’. These Communications announce inter alia an impact assessment on a European solution for orphan works.
the best models to maximise access and use of digitised material for the economy and society at large, in particular the fundamental conditions that should be respected in public-private partnerships for digitisation of works that are in the public domain;

- the role and responsibilities of private and public organisations for digitising orphan works as well as material that is in-copyright but no longer commercially available, with a view of tackling the risk of a ‘20th century black hole’ in Europeana and on the Internet in general;

- promoting the widest access to the digitised material across borders;

- ensuring sustainability of the digitised resources for long term preservation purposes.

The annex contains a list of more detailed questions in relation to these Terms of Reference.

Working Methods

The members of the Group will be appointed by Commissioners Kroes and Vassiliou. The Group will consist of three members, selected on the basis of their knowledge, experience and track record as visionary thinkers in the areas of culture and technology.

The members of the Group will conduct their mission under their sole responsibility. The Group can hold consultations, as appropriate, with members of the European Parliament, the competent authorities of Member States and other stakeholders.

The European Commission will ensure the secretariat of the Group. Where necessary, the Commission will make arrangements to bring in supplementary external technical expertise to support the work of the Group.

The Group will report back in writing to Commissioners Kroes and Vassiliou as often as appropriate. The report addressing the priorities set in this mandate will be finalised as soon as possible and no later than 8 months after the first meeting of the Group.

Reflection Group on bringing Europe’s cultural heritage

The present annex details a series of questions to be addressed as a starting point for the work of the Reflection Group.

**Sources of Funding for Digitisation**

What level of public funding would be necessary to bring to Europe’s cultural institutions collections online? What would be an appropriate balance between public and private funding for such digitisation? Should the use of European funding be considered to trigger large scale digitisation processes across the Union?

**Maximising the Impact of Digitisation of In-Public Domain Material for the Economy and Society**

What access and re-use models guarantee the highest impact of investments in digitisation for the economy and society at large? In particular, should public domain content
that is digitised with public and/or private resources be freely accessible and re-usable by all? Is any form of exclusive control on digitised public domain material by private partners acceptable in exchange for investments in digitisation? If so, what is the maximum time and other conditions (e.g. according to the types of material) that might qualify the terms of such exclusivity? Can a difference be made in access and use policies between the commercial use of the digitised public domain material and non-commercial use? What alternatives can be envisaged for exploitation models that restrict the access to and use of digitised public domain material (e.g. advertising)?

FACILITATING THE DIGITISATION OF ORPHAN WORKS AND OUT-OF-PRINT WORKS

Are there elements in the copyright framework or in the practical organisation of rights clearance that need particular attention in order to facilitate the large scale digitisation of orphan works and of works that are no longer commercially available? Who (cultural institutions, private sector digitisers, rights holders themselves) should be responsible for assuring their online access, and what conditions and means would enable them to fulfil this task? If public funding is used to pay for rights clearance for their digitisation and online accessibility, what mechanisms can be envisaged to arrive at a fair price for clearing the rights?

LONG-TERM PRESERVATION

How can the long term preservation of material that is digitised be guaranteed? What are the responsibilities of the private and public partners for ensuring the long term availability of this material in the context of a partnership?

CROSS BORDER ACCESS

Is any form of territorial discrimination between users (e.g. based on IP addresses) in accessing or using the digitised public domain material acceptable (e.g. free access for users in their own country, and access against payment for users from other EU or third countries)? What are the consequences of an extension of access and use on the costs for clearing rights?
Overview of the results of the online consultation
Results of the online consultation of the Comité des Sages on bringing Europe’s cultural heritage online

Total number of contributions to the consultation on bringing cultural heritage online (1258) as a percentage

Number of contributions excluding private citizens (586) as a percentage
Funding Digitisation

How should the digitisation of Europe’s cultural heritage (books, maps, paintings, museum objects, films, archival documents, sound) held by cultural institutions be financed?

Results from all participants

- Solely public funding: 31%
- Predominantly public funding: 66%
- Solely private funding: 3%
- Predominantly private funding: 0%

Results from participants excluding private citizens

- Solely public funding: 31%
- Predominantly public funding: 87%
- Solely private funding: 3%
- Predominantly private funding: 0%
If public funding is used for digitisation, what is the key justification?

For this question, 2.5% (31 of all participants) answered by using the option - 'other'. Half of them (15) expressed their agreement with at least one of the three given possibilities. 10 of the 15 participants were of the opinion that at least one of the three given reasons is acceptable as a key reason for the use of public funding in digitisation, while 5 of them said that all three given reasons are equally relevant to justify such a purpose. 5 participants evoked the concepts of preservation, culture, education and research.
Could EU funding catalyse large-scale digitisation projects in the Member States?

Results from all participants

- 75%: No, public funding for digitisation is in principle an issue for Member States or the individual institutions
- 13%: No, but EU funding should be available for digitisation projects with a real European scope and added value
- 9%: Yes, EU funding can make the difference and accelerate digitisation processes across Europe, therefore specific funding needs to be made available
- 3%: Yes, but the existing instruments at European level (e.g. the Structural Funds, Competitiveness and Innovation Programme) are sufficient

Results from participants excluding private citizens

- 83%: No, public funding for digitisation is in principle an issue for Member States or the individual institutions
- 9%: No, but EU funding should be available for digitisation projects with a real European scope and added value
- 6%: Yes, EU funding can make the difference and accelerate digitisation processes across Europe, therefore specific funding needs to be made available
- 2%: Yes, but the existing instruments at European level (e.g. the Structural Funds, Competitiveness and Innovation Programme) are sufficient
How can a sensible selection be made in the Member States of what needs to be digitised first with the limited funds available?

Results from all participants

Results from participants excluding private citizens

Cultural institutions should not spend too much thought on selection, it is the user who will decide what is interesting
Cultural institutions should start with public domain works, because it takes less time and money than clearing the rights for digitising in-copyright works
Cultural institutions should decide individually on the basis of their knowledge and know-how
Cultural institutions should decide, but they should be guided by a clear plan at national level to avoid duplication of effort
This should be decided top-down at national level, if not all individual institutions will go their own way
Other

For this question, 2.2% (27 participants) answered by using the option - 'other'. 13 participants used this open question to reinforce one of the given possibilities. 7 participants answered by combining at least two of the five given possibilities 4 participants suggested the creation of an international plan in order to select what needs to be digitised first within the Member States. 4 other participants expressed the wish to have guidelines, as well as international coordination and cooperation between organisations. Three of the participants saw digitisation as an instrument to be used first for the most vulnerable works in order to save the heritage.
How should Europeana - the central portal giving direct access to the digitised collections of Europe’s cultural institutions - be funded?

For this question, 13% (169 participants) answered by using the option - ‘other’. This option represents the second highest percentage (after the 72% who assigned the European funding to fund Europeana central portal). The majority of these respondents, 118 participants, combined the 2 first choices. From the other 51 answers, 30 associated at least 2 of the pre-established possibilities (although not always the same ones). Some additional ideas to fund the Europeana central portal were brought up such as: private funding; donations; the sponsoring of collections as well as public-private partnerships.
Access and use models

If the digitisation of public domain content (out of copyright works) is funded by public funds, what is the access and use model that maximises the impact of the investment?

Results from all participants

- 71%: The digitised content should be freely accessible and usable for non-commercial purposes and commercial use should be allowed against payment.
- 25%: The digitised content should be freely accessible and usable for all (including for commercial re-use).
- 4%: The cultural organisations who have digitised should be able to change users for access and use to recoup the investment.

Results from participants excluding private citizens

- 74%: The digitised content should be freely accessible and usable for non-commercial purposes and commercial use should be allowed against payment.
- 20%: The cultural organisations who have digitised should be able to change users for access and use to recoup the investment.
- 6%: The digitised content should be freely accessible and usable for all (including for commercial re-use).
Should collections from public institutions who sell access to public domain content digitised with public funding be included in Europeana (and be accessible to the end user against payment)?

Results from all participants

- 61%: Yes, there is no problem with that
- 22%: Yes, but only if the user gets access to the content in a reasonable resolution and can use it for non-commercial purposes such as schoolwork or research
- 17%: No, content digitised with public funding featuring in Europeana should be accessible and usable for all

Results from participants excluding private citizens

- 44%: Yes, there is no problem with that
- 29%: Yes, but only if the user gets access to the content in a reasonable resolution and can use it for non-commercial purposes such as schoolwork or research
- 27%: No, content digitised with public funding featuring in Europeana should be accessible and usable for all
Would you be willing to pay for accessing the content that you find through Europeana?

Results from all participants

- Yes: 14%
- Yes, but only for in-copyright material: 43%
- No: 43%

Results from participants excluding private citizens

- Yes: 31%
- Yes, but only for in-copyright material: 25%
- No: 44%
Public-private partnerships for digitisation

If the digitisation of public domain content (out of copyright works) is funded by private partners, is it acceptable that the public domain content is not freely accessible to the end user, because the private partner charges for access?

If the digitisation of public domain content (out of copyright works) is funded by private partners, is it acceptable that the public domain content is not freely available to the end user for a considerable amount of time because the private partner charges for access during that time?
If the digitisation of public domain content (out of copyright works) is funded by private partners, is it acceptable that the public domain content is only freely accessible to end users in one specific country, and not to those in other countries, because the private partner charges for access in those countries?

Results from all participants
- Yes: 19%
- No: 81%

Results from participants excluding private citizens
- Yes: 30%
- No: 70%
Is exclusivity on the exploitation of digitised public domain content acceptable in exchange for an investment by a private partner in digitisation?

Results from all participants

- 76%: Yes, there is no problem with exclusivity on the exploitation of public domain works from public institutions
- 11%: Yes, any type of exclusivity on the exploitation of public domain works from public institutions is acceptable, including those where the end user has to pay for the digitised content, as long as the exclusivity is limited in time
- 5%: Yes, but only models where access for the end user remains free are acceptable (e.g. models based on advertising). No time limits should apply.
- 4%: Yes, but only models where access for the end user remains free are acceptable (e.g. models based on advertising). Exclusivity should be limited in time.
- 4%: No, any form of exclusivity in relation with public domain works from public institutions should be avoided

Results from participants excluding private citizens

- 70%: Yes, there is no problem with exclusivity on the exploitation of public domain works from public institutions
- 13%: Yes, any type of exclusivity on the exploitation of public domain works from public institutions is acceptable, including those where the end user has to pay for the digitised content, as long as the exclusivity is limited in time
- 8%: Yes, but only models where access for the end user remains free are acceptable (e.g. models based on advertising). No time limits should apply.
- 5%: Yes, but only models where access for the end user remains free are acceptable (e.g. models based on advertising). Exclusivity should be limited in time.
- 4%: No, any form of exclusivity in relation with public domain works from public institutions should be avoided
If you think that exclusive agreements with private partners on the exploitation of digitised public domain works are acceptable, but should be limited in time, what is the maximum term of exclusivity that should be allowed?

<table>
<thead>
<tr>
<th>Term</th>
<th>Results from all participants</th>
<th>Results from participants excluding private citizens</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 2 years</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>&lt; 5 years</td>
<td>9%</td>
<td>9%</td>
</tr>
<tr>
<td>&lt; 10 years</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>&gt; 25 years</td>
<td>75%</td>
<td>68%</td>
</tr>
</tbody>
</table>

I think exclusivity should not be allowed
What exploitation model of digitised content would you consider best suited for public-private partnerships for public domain material from cultural institutions?

This question received the highest percentage for the option - ‘other’. 20.9% (262 participants) answered by expressing other ideas then the suggested ones. There are three main strands of ideas emerging from these replies. For 93 participants the exploitation model for digitised content best suited for public private partnerships of public domain content should be determined by the market.

80 participants took the occasion to comment on the exploitation of digitised public domain content. 34 respondents were against the exploitation based of this material, either by the public or by the private sector. 46 of them expressed their disagreement with any kind of payment to access cultural goods and only 11 of them would accept a possible payment when downloading and/or printing the digital content.

41 participants suggested that the exploitation model of digitised content should depend on the terms of the public private partnerships, and also on the type of works and on the type of contents.
Who should take responsibility for the long term preservation of the digital files resulting from public private partnerships for digitisation?

For this question, 5.3% (66 participants) answered by using the option - ‘other’. 19 participants supported the second choice, but added that the cultural institution should respond with both private and public funding (instead of public funding only). For 13 participants, the responsibility for the long term preservation in the digitization process should be left up to the parties involved as well as to the specific project/s concerned.
Digitisation of in-copyright works

Who should be responsible for the digitisation and subsequent online accessibility of orphan works?

Results from all participants

- Cultural institutions: 39%
- Private sector actors involved in the digitisation of content: 3%
- Cultural institutions and private sector actors involved in the digitisation of content: 56%
- Other: 2%

Results from participants excluding private citizens

- Cultural institutions: 44%
- Private sector actors involved in the digitisation of content: 4%
- Cultural institutions and private sector actors involved in the digitisation of content: 50%
- Other: 2%

For this question, 2.8% (35 participants) answered by using the option - ‘other’. 11 of these answers were chosen to reinforce the 3rd option.
Who should be responsible for the digitisation and subsequent online accessibility of works that are in copyright but out of print/out of distribution?

Results from all participants
- Cultural institutions, through licences: 53%
- Rights holders: 3%
- Other private sector actors involved in the digitisation of content, through licences: 10%
- Right holders in principle, but cultural institutions through licences if the right holders do not digitise themselves: 3%

Results from participants excluding private citizens
- Cultural institutions, through licences: 58%
- Rights holders: 25%
- Other private sector actors involved in the digitisation of content, through licences: 13%
- Right holders in principle, but cultural institutions through licences if the right holders do not digitise themselves: 4%
What issue would have the greatest effect on the conclusion of licences between right holders and cultural institutions for the large scale digitisation of works that are in-copyright but out of print/out of distribution?

For this question, 4.2% (53 participants) answered by using the option - ‘other’. 21 of these answers expressed agreement with at least 1 of the 5 given options. 10 respondents indicated they did not know enough about the topic to answer. 6 participants were of the opinion that right holders should decide how to exploit the works that are in-copyright but out of print.
Are there elements in the legislative copyright framework that would need particular attention in view of the large scale digitisation of works that are in-copyright but out of print/out of distribution?

For this question, 7.8% (98 participants) answered by using the option - ‘other’. More than half of them (51) answered indicating they are not well enough informed about this area. The majority of these answers were given by private citizens. 11 participants used this question to emphasise their agreement with the first choice, saying that no changes are needed in the legislative copyright framework for the large scale digitisation of in-copyright works but out of print/out of distribution.
Can public institutions who obtain a licence to digitise in-copyright content and make it available in their national territory (e.g. limited to internet addresses with a specific country code) be expected to get a licence covering the EU, even if this is more expensive?

![Pie charts showing distribution of opinions among participants.]

- **Results from all participants**:
  - Yes, this is an absolute must for the development of Europeana to which all MS have committed: 39%
  - Yes, but only if the price of the EU-wide licence is reasonable: 8%
  - No, national money should be used to grant national access: 8%
  - No, the negotiations on this with right holders are not likely to lead to results: 8%
  - Other: 37%

- **Results from participants excluding private citizens**:
  - Yes, this is an absolute must for the development of Europeana to which all MS have committed: 31%
  - Yes, but only if the price of the EU-wide licence is reasonable: 13%
  - No, national money should be used to grant national access: 6%
  - No, the negotiations on this with right holders are not likely to lead to results: 43%
  - Other: 7%

For this question, 8.2% (103 of all participants) answered by using the option 'other'. 13 participants answered by indicating they did not know. 17 participants indicated that the rights holders should decide on licences, either on a national or on the EU level. Some participants expressed reluctance against the idea of licences covering the whole of the EU, saying this type of licences is desirable but expressing concerns about making them mandatory. Some other respondents called for legislation at European level.
If public funding is granted to pay right holders for the digitisation of their works, what conditions should be imposed in exchange for this funding?

From the 1258 answers, 12.2% (153 of all participants) answered by using the option - ‘other’. For the great majority (97 participants) these conditions should be subject to negotiations and can vary from one situation to another.
339 out of the 1258 participants (27%) made a comment or suggestion.

59 participants, all of them private company or association of companies stressed the need to maintain the current copyright rules. In order to know the status of a work (in-copyright, out of commerce or orphan) prior due diligent search facilitated by the ARROW tool must be undertaken in the country of publication. For out of print or non-commercialised works, it should be left at the discretion of right holders to decide whether such works shall be made freely accessible. The author’s moral rights must be respected. Statutorily imposed cut-off dates are to be avoided, as some works will still be in-copyright.

46 participants underlined that digitisation should be understood as the supply of a global public good. Therefore the replies advocated re-use of the material available on Europeana for educational and non-commercial purpose, avoidance of the privatisation of public domain, as well as free and world wide accessibility on the digital content.

For the digitisation and online accessibility of in-copyright works, participants other than the rights holders suggested several options, in particular: to change current copyright law (in particular by expanding the exceptions for cultural institutions), voluntary agreements between right holders and cultural institutions, and collective licensing schemes as a flexible tool to enable mass-rights clearance.

Financing was addressed in several of the comments. The suggestions vary from the need for public funding for digitisation in general, to funding at European level (for Europeana and digitisation) and ideas for alternative funding models such as advertising. Several comments addressed public private partnerships and supported the implementation of such partnerships: ‘we are optimistic about public and private sectors working to deliver digitising solutions although some of the questions asked require a combination of solutions. A review of digitisation projects in the EU nation states will help gauge the sum of activity as a whole.’ Some replies are, however, more critical: ‘private partners cherry pick content and impose many and varied terms and conditions. Public funding is very important as it should allow full unbarred reuse of the digital item, and will lead to further creativity and innovation enriching our society and economy’.

A further range of suggestions concerned organisational issues and the spread of knowledge. Examples of the ideas mentioned are: the need to strengthen international collaboration, the importance of a transfer of knowledge and skills between large and small institutions; the establishment of special research and service institutions - centres of excellence for cultural transmission; and on the job training places for students to assist in and learn about the digitisation process.

Some of the replies indicated the digitisation and online accessibility of our cultural heritage need to be tackled with a sense of urgency. One participant wrote: ‘Agissez vite; ne nous retrouvons pas dans une “tour de Babel!”’
Several comments addressed the questionnaire itself. 39 participants regretted that it consisted mostly of closed questions and indicated that the limitation of 100 characters for open replies was too restrictive. The use of closed questions was seen by some respondents to be inadequate for debating such a complex series of issues. Other participants indicated that the questionnaire was too long and the language used too technical. Some suggested that the questionnaire should have been provided in all national languages instead of just English, French and German.

There were also comments welcoming the survey and more in general the consultation exercise. Some participants thanked the Reflection group for the initiative and for the possibility to participate and give their opinion on this subject.

Country participation to the questionnaire

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of replies (1258)</th>
<th>Country</th>
<th>Number of replies (1258)</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>475</td>
<td>Denmark</td>
<td>5</td>
</tr>
<tr>
<td>Germany</td>
<td>235</td>
<td>Norway</td>
<td>4</td>
</tr>
<tr>
<td>Belgium</td>
<td>120</td>
<td>Estonia</td>
<td>4</td>
</tr>
<tr>
<td>Poland</td>
<td>77</td>
<td>Turkey</td>
<td>4</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>57</td>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>Netherlands</td>
<td>55</td>
<td>Russia</td>
<td>3</td>
</tr>
<tr>
<td>Italy</td>
<td>53</td>
<td>Australia</td>
<td>3</td>
</tr>
<tr>
<td>Spain</td>
<td>27</td>
<td>Luxembourg</td>
<td>3</td>
</tr>
<tr>
<td>Ireland</td>
<td>12</td>
<td>Czech Republic</td>
<td>3</td>
</tr>
<tr>
<td>Austria</td>
<td>11</td>
<td>Iceland</td>
<td>3</td>
</tr>
<tr>
<td>Sweden</td>
<td>11</td>
<td>Lithuania</td>
<td>2</td>
</tr>
<tr>
<td>Portugal</td>
<td>10</td>
<td>Bulgaria</td>
<td>2</td>
</tr>
<tr>
<td>Finland</td>
<td>10</td>
<td>Croatia</td>
<td>2</td>
</tr>
<tr>
<td>Switzerland</td>
<td>10</td>
<td>Latvia</td>
<td>1</td>
</tr>
<tr>
<td>Romania</td>
<td>9</td>
<td>Israel</td>
<td>1</td>
</tr>
<tr>
<td>USA</td>
<td>7</td>
<td>Serbia</td>
<td>1</td>
</tr>
<tr>
<td>Hungary</td>
<td>7</td>
<td>Slovenia</td>
<td>1</td>
</tr>
<tr>
<td>Slovakia</td>
<td>6</td>
<td>Macedonia</td>
<td>1</td>
</tr>
<tr>
<td>Cyprus</td>
<td>6</td>
<td>Brasil</td>
<td>1</td>
</tr>
<tr>
<td>Greece</td>
<td>6</td>
<td>Albania</td>
<td>1</td>
</tr>
<tr>
<td>Malta</td>
<td>5</td>
<td>South Africa</td>
<td>1</td>
</tr>
</tbody>
</table>

42 countries participated in total to the online consultation.
Summary Record of the Public Hearing of the Comité des Sages on bringing Europe’s cultural heritage online
Summary of the public hearing

Summary Record of the Public Hearing of the Comité des Sages on bringing Europe’s cultural heritage online 28th October 2010, 10:00 – 17:00

Charlemagne building, conference room Sicco Mansholt, Brussels

<table>
<thead>
<tr>
<th>AGENDA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>10.05 - 10.10</strong>  Opening by the Chairman</td>
</tr>
<tr>
<td><strong>10.10 - 12.45</strong> Morning session</td>
</tr>
<tr>
<td>10.15 - 11.30  Europeana Foundation - Jill Cousins</td>
</tr>
<tr>
<td>Federation of European Publishers (FEP) and Syndicat National de l’Edition (SnE) - Liv Vaisberg, Emily Cleevelly, Catherine Blache</td>
</tr>
<tr>
<td>Ministry of Culture, France - Nicolas Georges, Sophie Verrier</td>
</tr>
<tr>
<td>European Broadcasting Union (EBU) and Zweites Deutsches Fernsehen (ZDF) - Nicola Frank, Renate Dörr</td>
</tr>
<tr>
<td>European Bureau of Library, Information and Documentation Associations (EBLIDA) - Harald von Hielmcrone</td>
</tr>
<tr>
<td>11.30 - 12.45  European Visual Artists (EVA) - Carol Streul</td>
</tr>
<tr>
<td>Google - Santiago de la Mora, Antoine Aubert</td>
</tr>
<tr>
<td>European Newspaper Publishers Association (ENPA) - Sophie Scrive</td>
</tr>
<tr>
<td>Orange France Telecom - Pierre Geslot, Benoît Chantoin</td>
</tr>
<tr>
<td>Association des Cinémathèque Européennes(ACE) - Thomas Christensen</td>
</tr>
<tr>
<td><strong>14.15 - 16.45</strong> Afternoon session</td>
</tr>
<tr>
<td>14.14 - 15.05  European Writers’ Council (EWC) - Myriam Diocaretz</td>
</tr>
<tr>
<td>Maison des Auteurs, Belgium - Frédéric Young</td>
</tr>
<tr>
<td>Communia, European Network on the Digital Public Domain - Juan Carlos de Martin</td>
</tr>
<tr>
<td>European Federation of Magazine Publishers (FAEP) - Catherine Starkie</td>
</tr>
</tbody>
</table>
### AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Organisation/Entity</th>
<th>Contact Person(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.05 - 15.55</td>
<td>The British Library</td>
<td>Ben White</td>
</tr>
<tr>
<td></td>
<td>Association of European Performers’ Organisations (AEPO - ARTIS)</td>
<td>Guenaëlle Collet, Nick Yule</td>
</tr>
<tr>
<td></td>
<td>Belgian Archives</td>
<td>Lucie Verachten</td>
</tr>
<tr>
<td></td>
<td>European Federation of the Picture Industry (CEPIC)</td>
<td>Sylvie Fodor</td>
</tr>
<tr>
<td>15.55 - 16.45</td>
<td>MediaLibraryOnLine, Italy</td>
<td>Guilio Blasi</td>
</tr>
<tr>
<td></td>
<td>Digital Heritage Netherlands (DEN) Foundation</td>
<td>Annelies van Nispen</td>
</tr>
<tr>
<td></td>
<td>European Federation of Journalists</td>
<td>Pamela Morinière, Yuklan Wong</td>
</tr>
<tr>
<td></td>
<td>Multimedia Library &amp; Engineering Bureau (IRCAM)</td>
<td>Michael Fingerhut</td>
</tr>
<tr>
<td>16.45 - 17.00</td>
<td>Closing remarks</td>
<td></td>
</tr>
</tbody>
</table>

**Mr Maurice Lévy (Chairman)** welcomed the participants and informed them about the organisation of the public hearing. In particular, he presented the composition and the mandate of the Comité. The hearing is part of a broader consultation process, including an online consultation which triggered 1258 replies. The stakeholders views gathered through this consultation process will feed into the reflection group's work which will be concluded at the end of the year with a final report including recommendations. The report will be presented to the Commission and made public.

23 organisations asked to be heard and were divided into five groups: two batches of five organisations in the morning session; one group of five and two groups of four in the afternoon. Each group was followed by questions posed by the Members to the speakers. The Chairman informed participants that after the hearing, the following documents would be published on the web pages of the Comité: the position papers submitted by participants; summary minutes and a video recording of the hearing.
Morning session

GROUP I

Europeana Foundation - Jill Cousins

Federation of European Publishers (FEP) and Syndicat National de l’Edition (SnE) - Liv Vaisberg, Emily Cleevely, Catherine Blache

Ministry of Culture, France - Nicolas Georges, Sophie Verrier

European Broadcasting Union (EBU) and Zweites Deutsches Fernsehen (ZDF) - Nicola Frank, Renate Dörr

European Bureau of Library, Information and Documentation Associations (EBLIDA) - Harald von Hiemcrone

EUROPEANA FOUNDATION
The Europeana Foundation is the governing body of the Europeana service, based on collaboration between museums, archives, audiovisual collections and libraries. Europeana is about connecting society with culture. After the launch of a prototype in 2008, a service is now offered where users can have integrated access to cultural content. Europeana is based on an open source code to allow reuse. Europeana also attaches a high importance to the public domain, through the adoption of a public domain charter and of the public domain mark recently introduced by Creative Commons.

Key points:
Europeana expressed the following needs to improve their activities in bringing cultural heritage online:

- To ensure sustainable financing for the Europeana service, which provides a high European added value with a relatively small budget (4-5 million € per year).
- To simplify the clearance of copyright for cultural heritage, in particular for digitising out-of-print/out-of distribution works and bringing them online.
- To ensure EU-wide cross-border access to cultural heritage over the Internet: there is a risk that national silos of information arise from the way in which certain public private-partnerships based on public domain information are arranged.
- To develop the use of digital cultural content in schools and other educational settings.
- To foster the interoperability of data, and in particular the adoption of open metadata.
FEDERATION OF EUROPEAN PUBLISHERS (FEP) AND
SYNDICAT NATIONAL DE L’ÉDITION (SNE)

FEP and SnE stressed that publishers support digitisation of cultural heritage and making it available, in particular on the basis of consensual arrangements between rightholders and cultural institutions.

Key points:
FEP and SnE recommend:
• To preserve the stability of the current copyright legislation and to enforce it.
• To ensure the possibility of publishers to get a return on investment.
• To find out the status of a work through due diligent search before digitising and making an orphan work available, and to support the mutual recognition within the EU of national orphan work solutions or negotiated licenses.
• To differentiate models of funding for digitisation on the basis of the copyright status: scanning of public domain works should be financed mainly with public funding, while for in-copyright works the involvement of rightholders is necessary.
• To ensure that digitisation and online accessibility for out-of-print works are always done in agreement with rightholders (various models of licensing can be envisaged, but prior authorisation should always be obtained).
• The ARROW system is the key to solving the orphan works issue.

MINISTRY OF CULTURE, FRANCE

The representative of the Ministry illustrated legislation in place and the state of the political debate concerning digitisation of cultural content. France has developed an important large-scale digitisation policy; since 2005, France has been very supportive of the perspective of a European digital library. Means to finance digitisation include the use of a ‘para-fiscal’ tax on reprographic devices, and more recently a national loan scheme (Grand Emprunt).

Key points:
• France will proceed and accelerate the digitisation process, online accessibility and long-term preservation based on a balanced legal framework.
• Digitisation of public domain works should be financed mainly with public funding, while the digitisation and exploitation of copyrighted works should be done by or in partnership with rightholders.
• Any new EU legislation on orphan or out of print works should allow every Member State to continue using its own solutions, which are better adapted to the national situation and take into account the negotiations with rights holders.
EUROPEAN BROADCASTING UNION (EBU) AND ZWEITES DEUTSCHES FERNSEHEN (ZDF)

EBU and ZDF stressed that audiovisual content held by broadcasters archives (estimate: 28 million hours, most in analogue format) need to be fully considered as a part of our cultural heritage.

Key points:
- 70% of 5 million hours of non-digitised media material of broadcasters (films, video and audio recordings) is at risk of being lost. The funding issue is vital to avoid this loss.
- Clearing the rights for online use of broadcasters’ archives would mean going back to some 3 million contracts. This cannot be done on an individual basis. Rights clearance for bringing the BBC archives online would involve 800 persons working full time for 3 years. A modern copyright clearing system adapted to the digital era is needed.
- EBU has put forward proposals to make copyright licensing future-proof and technology-neutral in ‘Modern Copyright for Digital Media’.

EUROPEAN BUREAU OF LIBRARY, INFORMATION AND DOCUMENTATION ASSOCIATIONS

Key points:
- Right clearance has become a major issue for libraries.
- On-site access to the digitised works on the premises of the libraries is not enough. A solution must be found by modifying the copyright legislation, either reinforcing the exceptions to copyright, or through extended collective licensing. Probably both are necessary.
- As for orphan works, in order to ensure a cross-border effect, a mechanism of mutual recognition by Member States of their solutions is needed.

EXCHANGE WITH THE SAGES - QUESTIONS AND ANSWERS

Responding to a specific question on the need of a reform to copyright legislation, FEP replied that in their opinion EU copyright legislation would not require any change, while in certain Member States, parts of national copyright laws may need to be changed and adapted to the digital era. On the same topic Europeana expressed the view that also EU copyright legislation would need important changes.

EBU replied to a question on the impact of extended collective licensing for audiovisual works: the combined effect of the ‘country-of-origin’ principle with ‘extended collective licensing’ could allow the usage of the broadcasters’ archives in digital services, it would reduce time, costs, resources for clearing rights and resolve the issue of orphan works.

EBLIDA reminded that extended collective licensing (ECL) is a well established system in Nordic countries, and that any initiative questioning the ECL would lead to opposition.
Europeana was questioned concerning the expression ‘national silos’, and provided examples of public-private partnerships for digitisation of public domain works which result in content not being accessible online outside a specific country.

Other questions posed to the speakers concerned public domain in the digital environment, the conditions for successful public-private partnerships, sponsoring, advertising, the costs of digital preservation, the criteria for selecting content to be digitised, and the role played by users in this process.

GROUP II

**European Visual Artists (EVA)** - Carol Streul

**Google** - Santiago de la Mora, Antoine Aubert

**European Newspaper Publishers Association (ENPA)** - Sophie Scrive

**Orange France Telecom** - Pierre Geslot, Benoît Chantoin

**Association des Cinémathèque Européennes (ACE)** - Thomas Christensen

EUROPEAN VISUAL ARTISTS (EVA)

EVA represents collecting societies for fine arts and photography in Europe. EVA societies license rights on behalf of the artists as a fiduciary.

**Key points:**
- Libraries that wish to digitize and make books and other printed materials available need prior authorization for the protected images included in the publications.
- EVA societies urge for remuneration for uses of orphan works in order to prevent that orphan works are used for unfair competition with other protected works.
- Within the ARROW PLUS project, EVA is studying the technical registration measures already existing on the market and new measures to be developed in order to prevent the future creation of orphan works.

GOOGLE

Google Books is fully part of the corporate mission, organising and providing access to the world’s knowledge. Google Books Library Project has scanned 15 million works covering more than 100 languages. These books come from 40 different libraries, of which 9 are European (only public domain works in Europe). Under the Google Books Partner Programme, agreements have been reached with 35,000 rightholders concerning in-copyright works.
Key points:
- Web access is a complement to existing business models, not a substitute.
- Through digitisation it is possible to keep books alive: 25-35% of the digitised public domain books in Google Books are accessed at least once a month, which shows the interest in this material.
- The way forward is based on partnership, and on new legislation providing solutions for orphan works and out-of-print works. Commercial re-use of orphan works should be possible.

European Newspaper Publishers Association (ENPA)

Key points:
- A good basis for cooperation and agreement between libraries and newspaper publishers exists in all Member States. This should not be challenged by potential changes of copyright legislation (introduction of new exceptions) or by imposing mandatory conditions. Licensing agreements publishers/libraries are essential to authorise use of content.
- Public funding allocated to libraries should not lead to unfair competition with publishers’ business models. Moreover, libraries should not get into subscription and advertising models for access.
- An orphan works definition in the newspaper sector is unclear. It is important to acknowledge the different sectors specificities (books, magazines, newspapers).
- The concept of out-of-print works is questionable in the newspaper sector (most newspapers will be out of print in a day).

Orange France Telecom

Key points:
- Cultural material can become a powerful driver of new services on the internet. An open and distributed model (no dominant positions) is necessary for the use of the material.
- Europe could take leadership on digital cultural heritage. The contribution of the EU as a driving force capable of uniting the stakeholders is essential to: create the foundations of a European digital market; harmonize the regulatory framework; develop models of sustainable public private investment and coordinate the various stakeholders.
- A legal deposit requirement for digitised material should be considered.
ASSOCIATION DES CINÉMATHÈQUE EUROPÉENNES

Key points:

• Cinema heritage can be estimated as being for 1/3 in-copyright, for another 1/3 in the public domain, and for the remaining 1/3 orphan works. Therefore the solution of the orphan works issue is vital for cinema archives and museums. European legislation on orphan works should cover all different types of material, including audiovisual.

• Copyright clearance of motion pictures is complex. If film archives are to facilitate rights clearance, either significant funds to do so, or clearer legislation with non-commercial fair use exceptions, is needed.

• Digital ‘business models’ should be a cultural investment. Public heritage institutions should continue to receive funds to preserve and present Europe’s heritage.

EXCHANGE WITH THE SAGES - QUESTIONS AND ANSWERS

Google was asked whether it would be possible to provide figures concerning the cost of digitisation. The reply was that this is an extremely expensive activity, but no concrete figures were provided. Google was also asked to clarify exclusivity clauses possibly present in its partnerships with libraries. Google denied the existence of exclusive arrangements.

ACE was questioned concerning the potential interest of the private sector in financing the digital preservation. The reply that so far there has not been a great interest of the private sector in preserving film heritage.

In reply to a specific question concerning how they intend to intervene in funding digitisation, Orange replied that their preference is to intervene through investment in partnership rather than with sponsoring. The potential of involvement of SMEs as providers of added value services based on re-use of digital content is also very high. An obstacle to long-term investment is the pressure by shareholders for immediate return.
Afternoon session

GROUP I

**European Writers’ Council (EWC)** - Myriam Diocaretz

**Maison des Auteurs, Belgium** - Frédéric Young

**Communia, European Network on the Digital Public Domain** - Juan Carlos de Martin

**European Federation of Magazine Publishers (FAEP)** - Catherine Starkie

**Conseil Régional d’Aquitaine - i2S** - Hervé le Guyader, Adrien Poly

**EUROPEAN WRITERS’ COUNCIL (EWC)**

*Key points:*

- A clearer distinction in policy papers between the digitisation for preservation and digitisation for online availability would allow stakeholders to take a clearer position.
- Support for Europeana; libraries should take the lead in digitisation; authors want their works to be widely available.
- Out of print works (OPW) and orphan works (OW) should be treated separately.
- EWC focused on OPW in particular arguing that the issue must be considered from the authors’ perspective, who is the original and eventual right holder. Dealing with the issue requires an analysis of the actual contractual agreements and provisions between author and publisher. In the majority of cases, the rights revert to the author/heirs in which case they should be free to negotiate new agreements on future editions.
- A standard definition of out of print works (OPW) is needed, as the conditions defining when a work is out of print vary according to publishing and distribution factors and national contexts.
- At present there are no specific licences to handle OPW, as the rights revert to the author. In light of mass digitisation of OPW by libraries and other digitising bodies, new solutions may be required.

**MAISON D’AUTEURS, BELGIUM**

*Key points:*

- The promotion of wide access to works under the banner of access to cultural heritage does not necessarily promote authors’ interests.
• There is a need to distinguish between digitisation of works and their accessibility. The authors welcome initiatives to digitise their creations for archiving and preservation purposes. This forms part of a ‘shared cultural heritage’.
• Three priorities for the authors: 1) make others recognise their professional value; 2) enforcement of copyrights (fair remuneration and moral rights); and 3) role of authors in adding value to cultural material (editorial role) in the online environment.
• The debate on accessibility raises a series of questions for the authors: 1) opposition to a new European legal license that can simplify the overall management of copyright and preference for collecting management schemes and balanced agreements; 2) How to ensure that copyrights are not absorbed by a monopoly of commercial giants; 3) support for ARROW project for solving the issues relating to OW.

COMMUNIA, EUROPEAN NETWORK ON THE DIGITAL PUBLIC DOMAIN
European thematic network on the digital public domain funded by EU (2007-2011); comprises 50 members including academic institutions, NGOs and umbrella organisations; issued Public Domain Manifesto (http://publicdomainmanifesto.org).

Key points:
• Digitisation should be financed predominantly by public funding, private partners should add value to this content and create new services and easy to access platforms. The involvement of private partners offers great advantages especially in terms of sustainability.
• Cultural institutions should start digitising public domain works. European funding can accelerate the digitisation process in Europe.
• Suggests legislative reforms abolishing or limiting territorial licensing within the EU.
• Digitised public domain content should be freely accessible for both commercial use/re-use and non-commercial purposes. Payment is only acceptable for in-copyright material.
• Exceptions in the European Copyright Directive are to be broadened.
• ‘User-generated’ digitisation (photos and videos of cultural heritage by EU citizens) to be exploited.

EUROPEAN FEDERATION OF MAGAZINE PUBLISHERS (FAEP)
FAEP represents the interests of 15,000 publishing companies, turning out more than 50,000 titles per year, selling 20 billion copies read by 360 million Europeans. The periodical press industry includes retail magazines, specialist magazines, business to business magazines as well as scientific and academic journals.

Key points:
• Copyrighted works can only be made available in respect of applicable laws and with the prior authorisation of rights holders.
For orphan works a diligent search is crucial. The Memorandum of Understanding agreed upon in the framework of the High Level Expert Group on Digital Libraries establishes the relevant guidelines. The creation of a comprehensive database of rights holders and publishing companies could help. Member States should be able to adopt their own solutions.

Voluntary agreements and public-private partnerships are the most suitable approach for the digitisation and exploitation of digitised content.

Important that publishers investing in digitization are able to monetise that content. If a private sector partner has invested in digitisation, they should have the possibility to recoup their investment, for example by time limited exclusivity for that digitised content.

Should an archive/library wish to provide online access to in-copyright content, it should do so through licensing arrangements, not interfering with the normal exploitation of the work by the right holders.

**Key points:**
The project is organized around four axes:

- Implement innovative economic models for digitised cultural heritage.
- Create an international Master programme for the development of cultural digital content.
- Provide collaborative governance and a sustainable economic model through the creation of an investment fund which will be supported by two thirds industrial partners and one third public organizations and financial partners.
- Develop strategic partnerships with enterprises in the social media in Silicon valley on projects relating to new technologies and applications in new multimedia uses for culture, education, games, information etc. Strategic partnership already set up with network Joint Venture, which brings together SMEs and higher education institutions involved in the development of digital economy.
EXCHANGE WITH THE SAGES - QUESTIONS AND ANSWERS

Responding to a question as to whether artists feel secure about being represented by collecting societies, EWC argued that, unfortunately, authors do not appear clearly under right holders. Collecting societies by definition do not represent authors unless this is clearly indicated in their mandate. So in certain areas, authors feel that they are not represented by collecting societies and prefer to 'represent themselves'.

EWC further argued that OPW may represent a dynamic new market. The e-book market is based on back catalogues of publishing houses, but unfortunately authors do not often reclaim their rights in this context. We would therefore need to create a database for OPW to identify the status of the work and the rights holders, along the lines of ARROW for OW. The Maison d'Auteurs, Belgium, argued that a specific licence for OW should not be introduced at European level.

Responding to a question, the Maison d'Auteurs, Belgium, argued in favour of the role of authors in the process of digitization, which should not be a tool of exclusive use by certain actors. Cultural heritage is a shared value and its use and exploitation should be based on mutual agreements between relevant stakeholders. This premise is found in a project initiated by the University of Liège and based on a partnership between publishers, public and academic institutions.

In reply, the Aquitaine region explained that there is a strong potential for SMEs in these different steps of the digitization chain, especially as regards specific competences required at different stages and complementarities of results. This is why the logic clustering SMEs and stakeholders of different profiles (ie. academic and public institutions, cultural institutions, creators) is very useful. Another area deserving attention is the training needs for professionals involved in digital economy and creative content on line.
GROUP II

The British Library - Ben White

Association of European Performers’ Organisations (AEPO - ARTIS) - Guenaëlle Collet, Nick Yule

Belgian Archives - Lucie Verachten

European Federation of the Picture Industry (CEPIC) - Sylvie Fodor

THE BRITISH LIBRARY

Key points:
- Support for public-private partnerships to the extent that they endorse the outcome of the High Level Expert Group on Digital Libraries and maintain a balance between public and private interests.
- Funding for digitisation should be mainly public and with open terms and conditions, thereby allowing to associate commercial entities.
- Regret for the knowledge gap between Europe and the US due to the different regimes for public domain works: for example, the oldest known in-copyright work in the British Library dates back to 1859. This means that even for works from far before 1900 the rights status needs to be checked systematically. Sensible and negotiated cut-off dates are necessary, also taking into account that the US already has such a cut-off date (1923), which guarantees a much wider access to knowledge in the US than in Europe. The situation in Europe clearly leads to a digital ‘black hole of the 20th century’. Call for the Comité des Sages to consider facilitating discussions between libraries and authors’ representatives on mutually agreeable historical cut-off dates for digitisation.
- Support for ARROW (British Library is a partner), which should become a ‘one-stop shop’ for clearing rights of works to be digitised.

ASSOCIATION OF EUROPEAN PERFORMERS’ ORGANISATIONS (AEPO - ARTIS)

Key points:
- Europeana should be the model for the future.
- A sectoral approach is needed in order to tackle effectively problems linked with the digitisation of creative content. This applies also to OW.
- A solution for mass digitisation may come through agreements involving all stakeholders, which are premised on the following principles: identification of rights holders, authorisation by rights holders for digitisation of their work, fair remuneration.
• Diligent search is crucial, especially in the case of OW. As an example, diligent search has helped identify the rights holders for many recordings, which would have otherwise been considered as orphan. Introducing an exception for OW would not be a viable solution for European performers.

BELGIAN ARCHIVES

**Key points:**

• It is physically impossible to digitise everything and make it available. It is therefore fundamental that Europeana and other initiatives are not limited to simply providing digitised items, but also include metadata that lead to non-digital information (catalogues of libraries, inventories of archives services).

• A digitised copy can never replace the original. This is why valorisation and preservation of both digitised items and its originals remain indispensible.

• Digitising the existing heritage is as important as guaranteeing long-term preservation of digital born objects (objects created in digital format). The preservation of digital both objects is a complex exercise, as the creation process does not take place in a controlled manner.

• Sustainable preservation goes beyond digitisation. We need to consider the lifespan of support for digitised material (ie. Cd, DVD) and the data formats used. Key elements to a lasting cultural heritage for the future: awareness of the volatility of the media used (digital photographs, digital films), on the one hand, and preparation of a handbook with good practices, on the other.

EUROPEAN FEDERATION OF THE PICTURE INDUSTRY (CEPIC)

CEPIC is an international federation of picture agencies and libraries. Members include over 1,000 stock photo and footage libraries, major news agencies, art galleries and museums in 20 European countries. These agencies and libraries produce content as copyright holders; collect and distribute distribution rights on behalf of the visual creators they represent; market this creative material, still and moving images, in their country and worldwide through a professional network for online publications, advertising, magazines and book publishing.

In this capacity, CEPIC defended the interests of the sector that may be put at stake through a future business model of Europeana.

**Key points:**

• Support for Europeana as a non profit organisation (Europeana not to ‘interfere’ with market players). Europeana and the Digital Libraries Initiative are important counterpoints to Google Books.

• Costs relating to digitisation are not limited to the digitisation process as such, but also involve costs relating to cataloguing activities (ie. creation of databases, identification and cap-
turing of material, indexing). The sustainability of digital assets is a major consideration in the preservation of cultural heritage. Criticism expressed that these aspects are not covered in the Terms of Reference of the CdS and the online questionnaire.

- Plans at national level, also linked with the Europeana project, should take into consideration that an extensive amount of material has already been digitised by cultural heritage institutions having trading companies with own commercial image libraries and that the commercial use of this material creates income for these institutions and funds further digitisation.

- In the event of Europeana allowing its material to be used for commercial purposes, re-use should not be free of charge, even if public funds supported digitisation. The Europeana project should not undermine healthy market activity by creating a free stock image library. Europeana may therefore endanger what already works on the ground.

- By the same token, any exclusivity clause provided to a private partner will adversely affect the market.

- Possible solutions to the issues raised above can be sought through partnerships between Europeana and private entities (ie. photo agencies, picture libraries) and through working with established commercial models of licensing lending their experience and support to Europeana’s service.

**EXCHANGE WITH THE SAGES - QUESTIONS AND ANSWERS**

Responding to a question as to whether cut-off dates should be combined with licence fees, the British Library explained that 1900 seems like a reasonable solution in light of their experience of the digitisation project with Microsoft. An electronic based search for right holders of 19th century works may be combined with a search through publisher and authors to identify the rights holders of in copyright works. Cut-off dates should be meshed with licences.

In response to a question on what business model is appropriate for the treatment of future OW, CEPIC argued that the best archives today dispose the best metadata and are commercially successful. This confirms that quality and technologically apt archiving systems are competitive commercially.

On a similar point, AEPO-ARTIS explained that the difficulties to identify authors, may be minimised if relevant information is clearly and correctly indicated at the moment of registration of an artistic production (ie. CD, DVD).

When asked to express views on the positions presented by CEPIC as regards a possible conflict between Europeana using its material for commercial purposes and existing business models for cultural content on line, Europeana replied that its structures are still new and totally tested. A business model is foreseen to be published by end 2010. Europeana only uses
metadata and provides access to material held by other entities rather than stocking up mate-
rial, something that it can afford.

In response, CEPIC clarified that there is no problem with Europeana acting as content ag-
gregator.

In reply to a question as to what are the priorities or criteria used to select which material to
digitise, Belgian Archives explained that selection is driven by demand by the public, con-
tacts with universities in the framework of scientific projects, or the degree of risk to which
certain categories of material are exposed.
GROUP III

**MediaLibraryOnLine, Italy** – Giulio Blasi

**Digital Heritage Netherlands (DEN) Foundation** – Annelies van Nispen

**European Federation of Journalists** – Pamela Morinière, Yuklan Wong

**Multimedia Library & Engineering Bureau (IRCAM)** – Michael Fingerhut

**MEDIALIBRARYONLINE, ITALY**

MLOL has been the first network of digital public libraries in Italy. The service is now distributed in 4 Italian regions to about 1,000 civic public libraries in about 1,000 cities. The MLOL portal acts as an aggregator and gives access to a growing collection of digital objects including music, video, audiobooks, ebooks, newspapers and periodicals, professional databases, learning objects and other digital resources (both licensed in-copyright and public domain content). Access to this content is free for library users with username/password authentication. The collection is composed of both commercial contents licensed from publishers/distributors and public domain objects indexed through metadata harvesting from various content providers.

**Key points:**
- The focus in Europe is on digitisation of cultural heritage and building academic digital libraries. A part of the public demand concerns movies, music, novels, essays, audiobooks of recent literature, newspapers and periodicals. In this field a very large part of the market is left to American and Asian companies.
- MLOL developed an economically sustainable model to bring in copyright digital content to library users. This model could be shared at the European level starting with Europeana project.

**DIGITAL HERITAGE NETHERLANDS (DEN) FOUNDATION**

The DEN Foundation is the Dutch national ICT knowledge centre for cultural heritage. DEN supports archives, museums and other heritage institutions to improve their digital strategies and services. DEN encourages the institutions to invest in open technology, to implement ICT-standards and to make use of other tools that contribute to sustainable information services.

**Key points:**
- Highlighted the challenge of sustainability of many services with digitised collections. Digitisation remains very much an ‘ad hoc’ thing.
- Special attention is required for digital-born art and culture.
Main challenge for heritage institutions is to offer digital information easy to use and re-use at educational, recreational or commercial level. Thus public funding is needed for digital services in order to boost the knowledge society, the economy and a vibrant cultural life.

The exchange of data through portals gets more complicated with growing numbers of partners. Standardisation and interoperability is essential for improving access through search engines and other web services.

Cultural heritage in the public domain should stay accessible to the public, even if it is digitised through public-private partnerships. Private entities can invest and add value to public domain material and still gain profit from it.

For in copyright content online, Creative Commons and the recently launched Public Domain Mark are useful instruments for cultural institutions to share material over the Internet. At the same time, adjusting the legal framework to the digital age is a necessity.

Stressed the importance of a legal exemption from educational and cultural accessibility of non-commercial cultural heritage.

EUROPEAN FEDERATION OF JOURNALISTS
The EFJ represents members in 34 countries and coordinates the activities of unions affiliated to the International Federation of Journalists in Europe, with a focus on developing and promoting common interests in the fields of social, economic, cultural and media policy. Its aim is to protect and defend freedom of expression and information as well as journalists’ human rights.

Key points:
- The shift to new sources of information sharing (ie. twitter, blogs) has resulted in a loss of protection for professional journalists and a lack of remuneration for the reproduction of their work online.
- Maintaining moral and economic rights is key, as well as developing online paid models.
- It is crucial to identify who owns the work in each case. Collecting societies have a key role to play. Journalists transfer the rights for first publication of their work only. Subsequent publications should be subject to agreement between collecting societies and publishers.
- We need to consider technical ways to avoid future orphan works and enforce moral rights (i.e. signed works).

MULTIMEDIA LIBRARY & ENGINEERING BUREAU (IRCAM)
IRCAM (Institut de Recherche et Coordination Acoustique/Musique) is a European institute for science about music and sound and avant-garde electro-acoustical music. Its multimedia library was established in 1996 and is organisationally linked with the Centre Pompidou in Paris.
**Key points:**
- Need to reinforce the music content available on Europeana.
- At present, Europeana aggregates metadata from other content providers and aggregators. In addition, it should analyse the content of text, images, video, audio and provide a full text search for contents as is the case with some digital libraries.
- Europeana could copy contents with different purposes: 1) in order to extract words, concepts and use the resulting analysis as an index to full search 2) provide a faster and alternative online user access (cache) to the contents if it is temporary unavailable from the original provider.
- **Pleased in favour of an exception for copying for non profit organisations.** For instance, Europeana should not renegotiate the right to copy and use the documents already placed online by the content provider, when their purpose is to make it available to the public.
- Music is part of cultural heritage but online access to it has many obstacles. Rights clearance is complex and we need to find creative solutions. A European solution on orphan works should also comprise the music sector. An exception should be envisaged; collecting societies should maintain an escrow amount to remunerate authors in case the rights holders appear.

**EXCHANGE WITH THE SAGES - QUESTIONS AND ANSWERS**

In replying to a question regarding their business model, **MediaLibraryOnLine** explained that they represent an independent consortium of libraries which buy digital content and focus on digital lending. The use of the platform is allowed through licences, but the platform has a non competitive approach to publishers and producers of content as they don’t pay to be on the platform. The platform acts as an aggregator of content on behalf of the consortium of participating libraries that provide financing, but contents are not accessible free of charge.

In replying to a relevant question, **DEN** explained that there is an initiative on mass digitization in the Netherlands, but it is still in its early stages. The Nordic model is currently being examined as a useful example. DEN further argued that there a legal exception for OW and OPW should be introduced at EU level.

The CdS expressed interest in the revolution brought by new technologies for certain categories of professionals, such as journalists, but wondered how legislation can adequately follow these rapid developments.

The public hearing was video recorded and streamed live on the web, from the Digital Libraries pages of Europa, the European Commission’s web service. The video and the position papers are available at

Annex

Results of the *Cap Gemini technical* audit of Europeana
Europeana: Technical Review Europeana Technology Platform – Capgemini December 13th 2010

Introduction: our Capgemini approach

WHAT WE DID AND DID NOT DO

What we DID do
The finding in this report are based on 3 different input sources:

1. A conference call with Jill Cousins - Director of Europeana Foundation and The European Library
2. Extensive desk research into the documents and website references delivered by Jill Cousins (Description of work ECP 2008 DILI 558001, Vancis hosting proposal, Vancis received hosting proposal) + a number of website references. For this the Capgemini evaluated this inputs using the standard Capgemini checklist (see annex 1)
3. Based on these inputs, an extensive interview was conducted with Jan Molendijk, Technical and Operations Director of Europeana

What we DID NOT do
1. Our standard Capgemini approach in these cases is to conduct a number of interviews/ workshops with relevant stakeholders, both on the business/client side as well as on the technology side. For budget reasons we limited ourselves to the two interviews, one with Jill Cousins and another one with Jan Molendijk
2. Our standard Capgemini approach is to conduct a number of on premise technical stress tests in order to acquire first hand management information about the systems performance under duress. In this instance we have taken the input from the two Vancis proposals and the verbal clarification of Jan Molendijk - Technical and Operations Director - on the performance of the systems as a basis for our assessment and have not conducted any hands on testing ourselves
3. We have not been able to look into the specifics of the software production process since no valid data were made available to us
Understanding the question

From our information, these are the major questions to be answered: the Comité would like to be certain that the technology on which Europeana has been developed can be relied upon for further developments and extensions. The Comité has the following questions in mind:

1. Is the technology used to develop European a state-of-the-art and up-to-date technology or is there a risk that it will be rapidly outdated by technological evolutions?
2. Is this technology able to deal with a dramatic increase in terms of the number of objects indexed? Is this technology able to deal with a steep increase in the number of unique visitors?
3. What would be your main recommendations in order to make sure that Europeana has the most efficient technology and organization in order to deliver its mission?

The volumes for the current system are calculated at 4,000 concurrent users and to allow for 1 million unique visitors per week. Since the launch of the prototype no marketing has been done towards end users. This will recommence now and we expect to build to around 500,000 unique visitors per month by April 2011.

Our findings and suggestions

Is the technology used to develop European a state-of-the-art and up-to-date technology or is there a risk that it will be rapidly outdated by technological evolutions?

Based on our study of the separate documents delivered to us and both the interviews conducted, it is our understanding that the Europeana project is based on a robust technology foundation that includes Open Standards, architecture, security. We suggest to look into the possibilities of cloud computing (see next page).

Is this technology able to deal with a dramatic increase in terms of the number of objects indexed? Is this technology able to deal with a steep increase in the number of unique visitors?

To answer this question we would rather have done our own hands on stress testing. Based on our desk research into the two Vancis documents and the interview with Mr. Molendijk, it is our perception that the system should be able to deal with 500,000 unique users per month as planned. Yet this is something that Capgemini would rather test hands on by our own standard testing protocols. We expect that there is room for Database Optimization, but we can only judge that after thorough testing and hands on investigation.

What would be your main recommendations in order to make sure that Europeana has the most efficient technology and organization in order to deliver its mission?

The absence of an overall Software Architecture Document (SAD) describing the current and the future overall architecture of the platform is a crucial limitation. Most ingredients appear
to be there. They need to be collected and written down into one coherent document that is accepted by both the Europeana technical team as well as the representatives of the business side.

The absence of Software Measurement Metrics - although we have not found data that would suggest the software development process is underperforming, the absence of solid benchmark data makes it quite difficult - if not impossible - to verify the overall quality of the software development process in terms of productivity.

In absence of an overall SAD, the (future) possibilities of cloud computing have not carefully been assessed. Our view is that cloud computing is an important overall IT trend that should be incorporated into Europeana’s future architecture and therefore described in the SAD already mentioned. Cloud is more that just virtualization (Infrastructure as a Service), it also encompasses Platform as a Service and Software as a Service. The latter two are even more relevant to Europeana than the IaaS dimension of cloud technology which appear to be the prevailing perception of cloudtechnology within the Europeana team. See the last slide with our suggestion for a ‘journey to the cloud’ workshop.

Annex

CHECKLIST QUESTIONS (1)

Software Architecture Document:
- How easy is it to upscale when data and usage increase?
- Analytics: what do we know about current performance and stress test results?
- How can future growth be simulated / predicted / reviewed?

Infrastructure:
- Is current set-up (with load balancers and firewall) prepared for easy upscaling apps/db servers?

Software / applications / business logic:
- Search caching: possible?
- API strategy: how will it impact performance?
- Software architecture set-up to minimise database performance?

Content & data:
- Image resizing / rendering – why and how? What is impact on server performance?
- Multi-linguality: how has it been set-up?
- Taxonomy: who owns this?
- Document oriented database – already analyzed?
User Experience:
• Next to API-strategy, can search be offloaded with fixed structure (fixed/cached queries)?
• Social: what is the roadmap for social media usage and UGC?

CHECKLIST QUESTIONS (2)

Questions
• What are current (known) architecture bottlenecks for technology platform upscale requirements?
• Current stress-test execution: what are the results and what are the test scenarios?
• Database: what are the ideas with replacing current db with a document oriented version?
• What are the ideas with the destination sites?
  – Guided navigation to decrease amount of queries?
  – Fixed queries?
• What will be the API strategy (how are musea involved?) and how is it planned with the roadmap for the destination sites? Are there any predictions for impact on site traffic?
• What is the roadmap for user generated content (we understand that some experimentation is done) and use of social?

TOPICS AS DISCUSSED ON 16-11 WITH JAN MOLENDIJK
• Productivity: software development productivity is not measured, but expected to be good. Learnings from factory approaches are welcome. The current model for Europeana is not a factory model, but more a collaborative approach between multiple project partners (musea, universities, TNO, etc.). Standards and frameworks for data enrichment and quality assurance are installed and working.
• Optimization on database server is running at the moment, possible performance improvement with document based model.
• Currently there is no SAD, but most ingredients are there. It has not been collected and written down into one document.
• Performance testing is executed on a regular basis, and new scenario’s can be tested thoroughly based on logriles.
• The API approach is used already with best practice examples, but also with examples that cause a lot of load. For this, guidelines are under development to optimize.
• Family sourcing is used for translations.
• Social monitoring is partly implemented by the Marcom team.
• Semantic search / natural languages search will be implemented in the next versions.
• Sugar CRM is used for CRM towards partners.
• There are doubts on a positive business case for cloud. Discussing the impact on a strategic level seems not effective at this stage. However, it might be good to discuss future possibilities from the content council perspective (discuss specific challenges and opportunities and how to solve them with all partners involved).
JOURNEY TO THE CLOUD ENGAGEMENT

Capgemini can help me think more about cloud.

I agree we should be doing more with cloud. I know where we go next.

This stuff really works – in my organisation.

I am seeing real business value – delivered.

Prime Sponsor Mtg

Full Sponsor and Event Design

Initial Discussions

Planning and Scheduling

Execute ASE

Mobilise PoV actions

Execute PoV

Business Exploitation

- 2-4 Weeks
- Identify Attendees
- Understand client aspirations and maturity

- 2 Days
- Agree Business Value
- Identify Roadmap
- Identify 3-4 ‘Proof of Value’ actions

- 2-4 Weeks
- Demonstrate Possibility

- 8-10 Weeks
- Prove Business Value

1-2 Consultants engaged over period

CG J2C Team and associated experts supplement consulting team

CG J2C Lead

CG ITS delivery units working with you
Annex

Report of the study into the cost of digitisation by the UK Collections Trust
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

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

**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

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

**TOTALS:**

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

### Archives

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

**TOTALS:**

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

### Audiovisual Archives

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated total number of hours of AUDIO in EU cultural institutions</td>
<td>10.81</td>
<td>Million hours</td>
</tr>
<tr>
<td>Estimated total number of hours of VIDEO in EU cultural institutions</td>
<td>12.14</td>
<td>Million hours</td>
</tr>
<tr>
<td>Estimated total number of hours of FILM in EU cultural institutions</td>
<td>1.04</td>
<td>Million hours</td>
</tr>
<tr>
<td>Estimated total number of hours of AV in EU cultural institutions</td>
<td>23.99</td>
<td>Million hours</td>
</tr>
<tr>
<td>Estimated mean cost of digitising AUDIO</td>
<td>0.67</td>
<td>€bn</td>
</tr>
<tr>
<td>Estimated mean cost of digitising VIDEO</td>
<td>5.36</td>
<td>€bn</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
<td>Units</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Estimated mean cost of digitising FILM</td>
<td>1.03</td>
<td>€bn</td>
</tr>
<tr>
<td><strong>TOTALS:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated total cost of digitising AV MATERIAL in the EU</td>
<td>4.94</td>
<td>€bn</td>
</tr>
<tr>
<td><strong>OVERALL TOTALS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated MEAN cost of digitising LIBRARY COLLECTIONS in the EU</td>
<td>19.77</td>
<td>€bn</td>
</tr>
<tr>
<td>Estimated total cost of digitising MUSEUM COLLECTIONS in the EU</td>
<td>38.73</td>
<td>€bn</td>
</tr>
<tr>
<td>Estimated total cost of digitising NATIONAL ARCHIVES in the EU</td>
<td>41.87</td>
<td>€bn</td>
</tr>
<tr>
<td>Estimated total cost of digitising AV COLLECTIONS in the EU</td>
<td>4.94</td>
<td>€bn</td>
</tr>
<tr>
<td>Estimated total cost of digitising CULTURAL MATERIAL in the EU*</td>
<td>105.31</td>
<td>€bn</td>
</tr>
</tbody>
</table>

* 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:

- DiCoMo\(^1\)
- NUMERIC\(^2\)
- 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 activities\(^4\):

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

---

\(^1\) Digitisation Cost Model (DiCoMo) - [http://cat.inist.fr/?aModele=afficheN&cpsidt=17182074](http://cat.inist.fr/?aModele=afficheN&cpsidt=17182074)

\(^2\) NUMERIC Digitisation Costs study - [http://www.numeric.ws/](http://www.numeric.ws/)

\(^3\) JISC Digitisation Costs Study - [http://www.jisc.ac.uk/media/documents/programmes/digitisation/digitisation-costs-full.pdf](http://www.jisc.ac.uk/media/documents/programmes/digitisation/digitisation-costs-full.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:

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

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:

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

<table>
<thead>
<tr>
<th>Object/items/materials in collections</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books (incl. ‘Rare Books’)</td>
<td>Published, printed books</td>
</tr>
<tr>
<td>Newspapers</td>
<td>Published, printed newspapers</td>
</tr>
<tr>
<td>Journals and other serials</td>
<td>Periodicals, magazines or other published material in a serialised form</td>
</tr>
<tr>
<td>Government publications</td>
<td>Printed material produced by Government or Government agencies</td>
</tr>
<tr>
<td>Other printed material</td>
<td>Other printed material not confirming to any of the above definitions</td>
</tr>
<tr>
<td>Manuscripts</td>
<td>Handwritten book or document</td>
</tr>
<tr>
<td>Maps</td>
<td>Map of geographical areas or features in a printed form</td>
</tr>
<tr>
<td>Photographs</td>
<td>Photographic reproductions, produced in hard copy</td>
</tr>
<tr>
<td>Microforms</td>
<td>Printed micro-reproductions of works such as newspapers</td>
</tr>
<tr>
<td>Engravings</td>
<td>A print made from an engraved plate</td>
</tr>
<tr>
<td>Drawings</td>
<td>An illustration drawn by hand in a hard copy form</td>
</tr>
<tr>
<td>Posters</td>
<td>Printed posters</td>
</tr>
<tr>
<td>Postcards</td>
<td>Printed postcard</td>
</tr>
<tr>
<td>Sheet music</td>
<td>A musical composition in printed or written form</td>
</tr>
<tr>
<td>Other images not listed above</td>
<td>Other forms of visual representation not included in the above list</td>
</tr>
<tr>
<td>Archived Government records</td>
<td>Information records produced by Government or Government agencies</td>
</tr>
<tr>
<td>Archived historic records</td>
<td>Information records of particular historical significance</td>
</tr>
<tr>
<td>All other archived records</td>
<td>And other form of information record not included above</td>
</tr>
<tr>
<td>Man-made artefacts in museums</td>
<td>Man-made objects, tools, artefacts included in museum collections</td>
</tr>
<tr>
<td>Natural specimens</td>
<td>Natural material (flora, fauna, mineral) in museum collections</td>
</tr>
<tr>
<td>2D works of art</td>
<td>Works of art in a 2D format, such as paintings, illustrations &amp; prints</td>
</tr>
<tr>
<td>3D works of art</td>
<td>Works of art in a 3D format, such as sculptures or other fabrications</td>
</tr>
<tr>
<td>Other museum objects</td>
<td>Other objects in collections of a type not confirming to the above list</td>
</tr>
<tr>
<td>Simple Classification</td>
<td>NUMERIC Classifications</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>Books (simple)</td>
<td>• Books</td>
</tr>
<tr>
<td>Books (complex)</td>
<td>• Rare Books</td>
</tr>
<tr>
<td>Simple 2D</td>
<td>• Journals</td>
</tr>
<tr>
<td></td>
<td>• Posters</td>
</tr>
<tr>
<td></td>
<td>• Postcards</td>
</tr>
<tr>
<td></td>
<td>• Archived Government records</td>
</tr>
<tr>
<td></td>
<td>• Other printed material</td>
</tr>
<tr>
<td></td>
<td>• Other archived material</td>
</tr>
<tr>
<td>Complex 2D</td>
<td>• Newspapers</td>
</tr>
<tr>
<td></td>
<td>• Manuscripts</td>
</tr>
<tr>
<td></td>
<td>• Maps</td>
</tr>
<tr>
<td></td>
<td>• Photographs</td>
</tr>
<tr>
<td></td>
<td>• Engravings</td>
</tr>
<tr>
<td></td>
<td>• Drawings</td>
</tr>
<tr>
<td></td>
<td>• Sheet Music</td>
</tr>
<tr>
<td></td>
<td>• Microforms</td>
</tr>
<tr>
<td></td>
<td>• 2D works of Art</td>
</tr>
<tr>
<td></td>
<td>• Archived historic records</td>
</tr>
<tr>
<td>Simple 3D</td>
<td>• Man-made artefacts (simple)</td>
</tr>
<tr>
<td></td>
<td>• Other museum objects</td>
</tr>
<tr>
<td>Complex 3D</td>
<td>• Man-made artefacts (complex/fragile)</td>
</tr>
<tr>
<td></td>
<td>• Natural history specimen</td>
</tr>
<tr>
<td></td>
<td>• 3D works of Art</td>
</tr>
<tr>
<td></td>
<td>• Monuments</td>
</tr>
<tr>
<td></td>
<td>• Landscape</td>
</tr>
<tr>
<td>Large-scale 3D</td>
<td>• Man-made artefacts (large)</td>
</tr>
<tr>
<td></td>
<td>• Natural history specimen (large)</td>
</tr>
<tr>
<td>Simple Audio</td>
<td>• Audio recordings (simple)</td>
</tr>
<tr>
<td>Complex Audio</td>
<td>• Audio recordings (multi-part, fragile)</td>
</tr>
<tr>
<td>Simple Video</td>
<td>• Video recordings (simple)</td>
</tr>
<tr>
<td></td>
<td>• Film footage</td>
</tr>
<tr>
<td></td>
<td>• Other time-based visual media</td>
</tr>
<tr>
<td>Complex Video</td>
<td>• Video recordings (complex, multi-part, fragile)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrastructure</td>
<td>Organisation type</td>
<td>Selection from a defined list</td>
</tr>
<tr>
<td></td>
<td>Organisation status</td>
<td>Multiple selection from a defined list indicating the legal/governance status of the institution.</td>
</tr>
<tr>
<td></td>
<td>Annual turnover</td>
<td>Annual operating budget for the institution</td>
</tr>
<tr>
<td></td>
<td>Dedicated Digitisation space</td>
<td>Availability of dedicated space for Digitisation</td>
</tr>
<tr>
<td></td>
<td>Dedicated staff</td>
<td>Availability of dedicated staff for Digitisation, including staff attached to partner organisations</td>
</tr>
<tr>
<td></td>
<td>Est. % Collection already Digitised</td>
<td>Estimation of the proportion of Collections already digitised (as an indicator of experience/expertise)</td>
</tr>
<tr>
<td></td>
<td>Dedicated legal support</td>
<td>Availability of dedicated legal advice and support (including capacity and experience in Rights Clearance)</td>
</tr>
<tr>
<td>Content</td>
<td>Books (simple)</td>
<td>As described above</td>
</tr>
<tr>
<td></td>
<td>Books (complex)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple 2D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex 2D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple 3D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex 3D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Large-scale 3D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple Audio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex Audio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simple Video</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex Video</td>
<td></td>
</tr>
<tr>
<td>Rights Status</td>
<td>Copyright</td>
<td>% of material for which copyright clearances required</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Orphan works</td>
<td>Proportion of material for which the copyright owner is either unknown or cannot be traced</td>
<td></td>
</tr>
<tr>
<td>Out of print</td>
<td>Proportion of material that is known to be out-of-print</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outputs</th>
<th>Hi-resolution image</th>
<th>Whether the Digitisation will produce hi-resolution, archival quality images suitable for preservation purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promotional image</td>
<td>Promotional image</td>
<td>Whether the Digitisation will produce attractive, specifically-lit images for use in promotional activity</td>
</tr>
<tr>
<td>Lo-resolution image</td>
<td>Lo-resolution image</td>
<td>Whether the Digitisation will produce lo-resolution, archival quality images suitable for online distribution</td>
</tr>
<tr>
<td>OCR output</td>
<td>OCR output</td>
<td>Whether the Digitisation workflow will use Optical Character Recognition (OCR) scanning</td>
</tr>
<tr>
<td>Encoded video</td>
<td>Encoded video</td>
<td>Whether Digitisation will produce encoded audio in an industry-standard file format</td>
</tr>
<tr>
<td>Encoded audio</td>
<td>Encoded audio</td>
<td>Whether Digitisation will produce encoded video</td>
</tr>
<tr>
<td>TEI XML</td>
<td>TEI XML</td>
<td>Whether the Digitisation will produce text with structured markup in Text Encoding Initiative (TEI) XML format</td>
</tr>
<tr>
<td>Post-production</td>
<td>Colour correction/balancing</td>
<td>Whether colour correction will be done using photo management software in post-production</td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>Quality Assurance</td>
<td>Whether resources are allocated to quality management</td>
</tr>
<tr>
<td>Metadata creation</td>
<td>Metadata creation</td>
<td>Whether the Digitisation will produce new metadata records about the content</td>
</tr>
<tr>
<td>Metadata enrichment</td>
<td>Metadata enrichment</td>
<td>Whether the Digitisation workflow will enrich existing metadata records about the content</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of potential options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bound/unbound</td>
<td>2</td>
</tr>
<tr>
<td>Page size (8.5x11, 11x17, 17x22)</td>
<td>3</td>
</tr>
<tr>
<td>Scanning resolution (300dpi, 400dpi, 600dpi)</td>
<td>3</td>
</tr>
<tr>
<td>Scanning bit depth (1, 8, 24)</td>
<td>3</td>
</tr>
<tr>
<td>Handling (fragile/non-fragile)</td>
<td>2</td>
</tr>
<tr>
<td>Place of performance (on/off site)</td>
<td>2</td>
</tr>
</tbody>
</table>

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):

<table>
<thead>
<tr>
<th>Object/items/materials in collections</th>
<th>Units</th>
<th>Median results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Digitised as %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of collection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pages per</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost per</td>
</tr>
<tr>
<td></td>
<td></td>
<td>unit (€)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cost per</td>
</tr>
<tr>
<td></td>
<td></td>
<td>page (€)</td>
</tr>
<tr>
<td>Books (incl. ‘Rare Books’)</td>
<td>Volumes</td>
<td>0.05</td>
</tr>
<tr>
<td>Newspapers</td>
<td>Volumes</td>
<td>1.7</td>
</tr>
<tr>
<td>Journals and other serials</td>
<td>Volumes</td>
<td>0.1</td>
</tr>
<tr>
<td>Government publications</td>
<td>Volumes</td>
<td>16</td>
</tr>
<tr>
<td>Other printed material</td>
<td>Number</td>
<td>5.9</td>
</tr>
<tr>
<td>Manuscripts</td>
<td>Number</td>
<td>0.3</td>
</tr>
<tr>
<td>Maps</td>
<td>Number</td>
<td>1.2</td>
</tr>
<tr>
<td>Photographs</td>
<td>Number</td>
<td>3.9</td>
</tr>
<tr>
<td>Engravings</td>
<td>Number</td>
<td>4.4</td>
</tr>
<tr>
<td>Drawings</td>
<td>Number</td>
<td>20</td>
</tr>
<tr>
<td>Posters</td>
<td>Number</td>
<td>18</td>
</tr>
<tr>
<td>Postcards</td>
<td>Number</td>
<td>33</td>
</tr>
<tr>
<td>Sheet music</td>
<td>Number</td>
<td>1.3</td>
</tr>
<tr>
<td>Other images not listed above</td>
<td>Number</td>
<td>18</td>
</tr>
<tr>
<td>Archived Government records</td>
<td>Metres</td>
<td>50</td>
</tr>
<tr>
<td>Archived historic records</td>
<td>Metres</td>
<td>25</td>
</tr>
<tr>
<td>All other archived records</td>
<td>Metres</td>
<td>54</td>
</tr>
<tr>
<td>Man-made artefacts in museums</td>
<td>Artefacts</td>
<td>53</td>
</tr>
<tr>
<td>Natural specimens</td>
<td>Objects</td>
<td>-</td>
</tr>
<tr>
<td>2D works of art</td>
<td>Exhibits</td>
<td>72</td>
</tr>
<tr>
<td>3D works of art</td>
<td>Exhibits</td>
<td>75</td>
</tr>
<tr>
<td>Other museum objects</td>
<td>Objects</td>
<td>10</td>
</tr>
<tr>
<td>Film and video recordings</td>
<td>Hours</td>
<td>3.4</td>
</tr>
<tr>
<td>Music and recorded sound</td>
<td>Hours</td>
<td>9.1</td>
</tr>
<tr>
<td>Other items not listed above</td>
<td>Number</td>
<td>45</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Institution</th>
<th>No need to digitise</th>
<th>Digitisation completed</th>
<th>Awaiting digitisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archives</td>
<td>36%</td>
<td>1%</td>
<td>63%</td>
</tr>
<tr>
<td>Broadcasters</td>
<td>28%</td>
<td>6%</td>
<td>66%</td>
</tr>
<tr>
<td>Museums</td>
<td>3%</td>
<td>25%</td>
<td>72%</td>
</tr>
<tr>
<td>Libraries</td>
<td>69%</td>
<td>1%</td>
<td>30%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
<td>15%</td>
<td>85%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>31%</td>
<td>11%</td>
<td>58%</td>
</tr>
</tbody>
</table>
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 population\(^6\).

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 2000\(^7\), 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.

<table>
<thead>
<tr>
<th>Media</th>
<th>Quantity (estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>5.64 billion books, 1.41 trillion pages</td>
</tr>
<tr>
<td>Microform</td>
<td>0.02 billion units</td>
</tr>
<tr>
<td>AV</td>
<td>Up to 6.54 billion units</td>
</tr>
<tr>
<td>Other</td>
<td>Up to 3.69 billion units</td>
</tr>
<tr>
<td>TOTALS</td>
<td>15.89 billion units (+/- 15%)</td>
</tr>
</tbody>
</table>

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.

7.6.5 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:

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

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:

<table>
<thead>
<tr>
<th>Description</th>
<th>Multiplier</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting figure</td>
<td>-</td>
<td>191.35m</td>
</tr>
<tr>
<td>Already digitised</td>
<td>1%</td>
<td>189.44</td>
</tr>
<tr>
<td>If 69% not eligible for Digitisation</td>
<td>189.44m less 69%</td>
<td>58.73m</td>
</tr>
<tr>
<td>If 50% not eligible for Digitisation</td>
<td>189.44m less 50%</td>
<td>94.72m</td>
</tr>
<tr>
<td>Upper estimate for total number of pages</td>
<td></td>
<td>2.36 bn pages</td>
</tr>
<tr>
<td>Lower estimate for total number of pages</td>
<td></td>
<td>1.47 bn pages</td>
</tr>
</tbody>
</table>
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).

This means that the total number of *pages eligible* for Digitisation ranges from 1.47bn up to 2.36bn.

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.

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.

### How Much Per Page?

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.

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.

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).

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.**

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Per-volume Cost (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital costs (per volume)</td>
<td>4.50</td>
</tr>
<tr>
<td>Preparation, selection and unbinding</td>
<td>4.50</td>
</tr>
<tr>
<td>(plus conditional assessment)</td>
<td></td>
</tr>
<tr>
<td>Scanning costs (£0.15 per page x 250</td>
<td>37.5</td>
</tr>
<tr>
<td>pages)</td>
<td></td>
</tr>
<tr>
<td>OCR conversion costs (£0.08 per page</td>
<td>20</td>
</tr>
<tr>
<td>x 250 pages)</td>
<td></td>
</tr>
<tr>
<td>PDF conversion costs (£0.2 per page</td>
<td>50</td>
</tr>
<tr>
<td>x 250 pages)</td>
<td></td>
</tr>
<tr>
<td>Simple metadata creation (per volume,</td>
<td>7.5</td>
</tr>
<tr>
<td>simple DC record)</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL per volume</strong></td>
<td><strong>€124 per book</strong></td>
</tr>
</tbody>
</table>
High estimate:

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

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:

<table>
<thead>
<tr>
<th>Books (lower estimate)</th>
<th>Books (upper estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost per book (lower estimate)</td>
<td>€7.32bn</td>
</tr>
<tr>
<td>Cost per book (upper estimate)</td>
<td>€10.06bn</td>
</tr>
</tbody>
</table>

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) \( \times 250 \)
- OCR conversion costs (per page) \( \times 250 \)
- PDF conversion costs (per page) \( \times 250 \)
- 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*

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

**TOTAL per volume** €73.15 per book

*High estimate:*

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

**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:

<table>
<thead>
<tr>
<th>Books (lower estimate)</th>
<th>Books (upper estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost per book (lower estimate)</strong></td>
<td><strong>€4.32bn</strong></td>
</tr>
<tr>
<td><strong>Cost per book (upper estimate)</strong></td>
<td><strong>€6.95bn</strong></td>
</tr>
<tr>
<td><strong>Cost per book (lower estimate)</strong></td>
<td><strong>€6.11bn</strong></td>
</tr>
<tr>
<td><strong>Cost per book (upper estimate)</strong></td>
<td><strong>€9.83bn</strong></td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Item</th>
<th>Per-volume Cost (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Management</td>
<td>2.15</td>
</tr>
<tr>
<td>Preparation, selection and unbinding (plus conditional assessment)</td>
<td>4.50</td>
</tr>
<tr>
<td>Scanning costs (€0.08 per page x 250 pages)</td>
<td>20</td>
</tr>
<tr>
<td>OCR conversion costs (€0.03 per page x 250 pages)</td>
<td>7.50</td>
</tr>
<tr>
<td>PDF conversion costs (€0.07 per page x 250 pages)</td>
<td>17.50</td>
</tr>
</tbody>
</table>
Simple metadata creation (per volume, simple DC record) 3.50

TOTAL per volume €55.15 per book

High estimate:

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

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:

<table>
<thead>
<tr>
<th></th>
<th>Books (lower estimate)</th>
<th>Books (upper estimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost per book (lower estimate)</td>
<td>€3.25bn</td>
<td>€5.24bn</td>
</tr>
<tr>
<td>Cost per book (upper estimate)</td>
<td>€3.92bn</td>
<td>€6.32bn</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th></th>
<th>Low estimate (€bn)</th>
<th>High estimate (€bn)</th>
<th>Mean (€bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-house, project-based Digitisation</td>
<td>7.32</td>
<td>16.20</td>
<td>11.76</td>
</tr>
<tr>
<td>Oursourced, project-based Digitisation</td>
<td>4.32</td>
<td>9.83</td>
<td>7.08</td>
</tr>
<tr>
<td>Public-Private Partnership Digitisation</td>
<td>3.25</td>
<td>6.32</td>
<td>4.79</td>
</tr>
</tbody>
</table>
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* 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:

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

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,


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 submission 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 indicated 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.ifrro.org/content/ifrro-recommends-guidelines-facilitate-solutions-licensing-out-print-works
<table>
<thead>
<tr>
<th>National Libraries</th>
<th>5,678,632</th>
<th>154</th>
<th>0.87</th>
</tr>
</thead>
<tbody>
<tr>
<td>University Libraries</td>
<td>3,796,669</td>
<td>117</td>
<td>0.44</td>
</tr>
<tr>
<td>Public Libraries</td>
<td>12,735,200</td>
<td>122</td>
<td>1.55</td>
</tr>
<tr>
<td>Special Libraries</td>
<td>873,406</td>
<td>426</td>
<td>0.37</td>
</tr>
</tbody>
</table>

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\textsuperscript{11} provides the following useful chart of estimated costs for different Rare Book activities in libraries:

![Cost Chart](http://www.clir.org/pubs/reports/pub103/appendix6.html)

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:

\textsuperscript{11}http://www.clir.org/pubs/reports/pub103/appendix6.html
| 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:

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

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. The

---

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:
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

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).

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

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):
<table>
<thead>
<tr>
<th>Material type</th>
<th>Estimate</th>
<th>Per-image cost (€)</th>
<th>Total estimated cost (€m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Low</td>
<td>4</td>
<td>115.2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>6</td>
<td>172.8</td>
</tr>
<tr>
<td>Complex</td>
<td>Low</td>
<td>9</td>
<td>259.2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>12</td>
<td>345.6</td>
</tr>
<tr>
<td>Oversized</td>
<td>Low</td>
<td>12</td>
<td>345.6</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>15</td>
<td>432</td>
</tr>
</tbody>
</table>

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):

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

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:

<table>
<thead>
<tr>
<th>Material</th>
<th>Type of Digitisation</th>
<th>Lower estimate</th>
<th>Higher estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books</td>
<td>In-house</td>
<td>7.32</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>Outsourced</td>
<td>4.32</td>
<td>9.83</td>
</tr>
<tr>
<td></td>
<td>Public Private Partnership</td>
<td>3.25</td>
<td>6.32</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Category</th>
<th>Lower estimate (€bn)</th>
<th>Higher estimate (€bn)</th>
<th>MEAN estimate (€bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare books (incl. manuscripts &amp; incunabula)</td>
<td>6.73</td>
<td>10.51</td>
<td></td>
</tr>
<tr>
<td>Archives</td>
<td>1.85</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td>Newspapers</td>
<td>0.52</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Maps</td>
<td>0.018</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>Photographs</td>
<td>0.014</td>
<td>0.19</td>
<td></td>
</tr>
</tbody>
</table>

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 session:\(^\text{13}\):

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 2008\(^\text{14}\). 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 museum\(^\text{15}\):

---

\(^{13}\) 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 (KDCS) 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.

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

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.

---

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:

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

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)\(^\text{17}\):

<table>
<thead>
<tr>
<th>Cost centre</th>
<th>Actual cost (€)</th>
<th>% of total costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff costs</td>
<td>80500</td>
<td>44%</td>
</tr>
<tr>
<td>Overheads</td>
<td>14559</td>
<td>8%</td>
</tr>
<tr>
<td>Equipment &amp; software</td>
<td>86961</td>
<td>47%</td>
</tr>
<tr>
<td>TOTAL cost</td>
<td>182021</td>
<td></td>
</tr>
<tr>
<td>Number of objects digitised</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>Cost per object</td>
<td>€121</td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{17}\) 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:

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

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:

<table>
<thead>
<tr>
<th>Art/Archaeology (m)</th>
<th>Science/Technology (m)</th>
<th>Other (m)</th>
<th>TOTAL (million objects)</th>
<th>Equivalent to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paintings</td>
<td>5.83</td>
<td>0.67</td>
<td>1.11</td>
<td>7.62</td>
</tr>
<tr>
<td>Sculptures</td>
<td>6.22</td>
<td>0.94</td>
<td>2.54</td>
<td>9.71</td>
</tr>
<tr>
<td>Posters/postcards</td>
<td>6.61</td>
<td>9.16</td>
<td>2.86</td>
<td>18.63</td>
</tr>
<tr>
<td>Drawings</td>
<td>7.37</td>
<td>20.15</td>
<td>1.44</td>
<td>28.96</td>
</tr>
<tr>
<td>Engravings/prints</td>
<td>7.08</td>
<td>1.52</td>
<td>1.91</td>
<td>10.51</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Museum Type</th>
<th>Number of photographs (median)</th>
<th>Number of museums</th>
<th>Estimated number of objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art/Archaeology/History Museums</td>
<td>10,000</td>
<td>7776</td>
<td>77.76m</td>
</tr>
<tr>
<td>Science/Technology Museums</td>
<td>53,600</td>
<td>6716</td>
<td>359.98m</td>
</tr>
<tr>
<td>Other types of Museum</td>
<td>15000</td>
<td>3181</td>
<td>47.72m</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>485.45m</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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:

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

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:

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

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

<table>
<thead>
<tr>
<th></th>
<th>Lower estimate (€bn)</th>
<th>Higher estimate (€bn)</th>
<th>Mean (€bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digitising museum collections</td>
<td>13.75</td>
<td>63.7</td>
<td>38.73</td>
</tr>
</tbody>
</table>

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.

---


19 Published by the Council of Europe – [http://consilium.europa.eu](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, including:

- 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

---

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 difficult 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:

<table>
<thead>
<tr>
<th>Unit</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pages per linear metre (historical archives)</td>
<td>5991</td>
</tr>
<tr>
<td>Pages per linear metre (Government archives)</td>
<td>7000</td>
</tr>
</tbody>
</table>

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):

<table>
<thead>
<tr>
<th>Archive type</th>
<th>Proportion already digitised (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government records</td>
<td>50</td>
</tr>
<tr>
<td>Historic records</td>
<td>25</td>
</tr>
<tr>
<td>Other types of archival material</td>
<td>54</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Countries</th>
<th>National Archives (linear metres)$^{21}$</th>
<th>BnPages (based on 6000 pages/metre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>177700</td>
<td>1.07</td>
</tr>
<tr>
<td>Belgium</td>
<td>200000</td>
<td>1.20</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>3000</td>
<td>0.02</td>
</tr>
<tr>
<td>Cyprus</td>
<td>4453</td>
<td>0.03</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>95000</td>
<td>0.57</td>
</tr>
<tr>
<td>Denmark</td>
<td>140000</td>
<td>0.84</td>
</tr>
<tr>
<td>Estonia</td>
<td>73000</td>
<td>0.44</td>
</tr>
<tr>
<td>Finland</td>
<td>180000</td>
<td>1.08</td>
</tr>
<tr>
<td>France</td>
<td>364000</td>
<td>2.18</td>
</tr>
</tbody>
</table>

$^{21}$ Based on reported figures from each National Archive (dates 2005-10)
<table>
<thead>
<tr>
<th>Country</th>
<th>Pages</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>300000</td>
<td>1.80</td>
</tr>
<tr>
<td>Greece</td>
<td>8000</td>
<td>0.05</td>
</tr>
<tr>
<td>Hungary</td>
<td>71000</td>
<td>0.43</td>
</tr>
<tr>
<td>Ireland</td>
<td>Data unavailable</td>
<td>0.00</td>
</tr>
<tr>
<td>Italy</td>
<td>160000</td>
<td>9.60</td>
</tr>
<tr>
<td>Latvia</td>
<td>13500</td>
<td>0.08</td>
</tr>
<tr>
<td>Lithuania</td>
<td>67500</td>
<td>0.41</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>32000</td>
<td>0.19</td>
</tr>
<tr>
<td>Malta</td>
<td>10000</td>
<td>0.06</td>
</tr>
<tr>
<td>Netherlands</td>
<td>93000</td>
<td>0.56</td>
</tr>
<tr>
<td>Poland</td>
<td>220000</td>
<td>1.32</td>
</tr>
<tr>
<td>Portugal</td>
<td>6000</td>
<td>0.04</td>
</tr>
<tr>
<td>Romania</td>
<td>Data unavailable</td>
<td>0.00</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>157000</td>
<td>0.94</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1300</td>
<td>0.01</td>
</tr>
<tr>
<td>Spain</td>
<td>220000</td>
<td>1.32</td>
</tr>
<tr>
<td>Sweden</td>
<td>300000</td>
<td>1.80</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>160000</td>
<td>0.96</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Estimated number of pages</th>
<th>Total estimated cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.94bn</td>
<td>€31.77bn</td>
</tr>
<tr>
<td>12.95bn</td>
<td>€51.80bn</td>
</tr>
</tbody>
</table>

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:

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

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:

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

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

<table>
<thead>
<tr>
<th>Material</th>
<th>Median cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival records</td>
<td>€41.79bn</td>
</tr>
<tr>
<td>Photographic collections in archives</td>
<td>€64.51m</td>
</tr>
<tr>
<td>Microform collections in archives</td>
<td>€11.73m</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>€41.87bn</strong></td>
</tr>
</tbody>
</table>
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*:

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

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

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

* 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 TAPE\(^ {22} \) 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](http://www.tape-online.net/survey.html)
<table>
<thead>
<tr>
<th>Format</th>
<th>Number of respondents</th>
<th>Total (x1000 hours)</th>
<th>Average (x 1000 hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>288</td>
<td>9386</td>
<td>33</td>
</tr>
<tr>
<td>Video</td>
<td>274</td>
<td>10559</td>
<td>39</td>
</tr>
<tr>
<td>Film</td>
<td>152</td>
<td>894</td>
<td>6</td>
</tr>
<tr>
<td>TOTALS</td>
<td>-</td>
<td>20839</td>
<td>-</td>
</tr>
</tbody>
</table>

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:

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

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:

<table>
<thead>
<tr>
<th>Format</th>
<th>Total (hours)</th>
<th>Extrapolated (+15%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>9.4m</td>
<td>10.81m</td>
</tr>
<tr>
<td>Video</td>
<td>10.56m</td>
<td>12.14m</td>
</tr>
<tr>
<td>Film</td>
<td>0.9m</td>
<td>1.04m</td>
</tr>
<tr>
<td>TOTALS</td>
<td>20.8m</td>
<td>29.92</td>
</tr>
</tbody>
</table>

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*

<table>
<thead>
<tr>
<th>Item</th>
<th>Per-hour Cost (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Management</td>
<td>3.50</td>
</tr>
<tr>
<td>Preparation &amp; movement</td>
<td>4.50</td>
</tr>
<tr>
<td>Preservation</td>
<td>10.00</td>
</tr>
<tr>
<td>Conversion costs</td>
<td>25.00</td>
</tr>
<tr>
<td>Rich metadata enhancement</td>
<td>7.00</td>
</tr>
<tr>
<td>TOTAL per volume</td>
<td>€50.00 per hour</td>
</tr>
</tbody>
</table>
High-end Audio Digitisation

<table>
<thead>
<tr>
<th>Item</th>
<th>Per-hour Cost (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Management</td>
<td>4.50</td>
</tr>
<tr>
<td>Preparation &amp; movement</td>
<td>5.00</td>
</tr>
<tr>
<td>Preservation</td>
<td>13.00</td>
</tr>
<tr>
<td>Conversion costs</td>
<td>36.50</td>
</tr>
<tr>
<td>Rich metadata enhancement</td>
<td>15.00</td>
</tr>
<tr>
<td><strong>TOTAL per volume</strong></td>
<td><strong>€74.00 per hour</strong></td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Item</th>
<th>Total Cost (€bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum estimate (based on 10.81m hours @ €50.00 per hour)</td>
<td>0.54</td>
</tr>
<tr>
<td>Maximum estimate (based on 10.81m hours @ €74.00 per hour)</td>
<td>0.80</td>
</tr>
<tr>
<td>Mean</td>
<td>0.67bn</td>
</tr>
</tbody>
</table>

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:

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Total Cost (€bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum estimate (based on 12.14m hours @ €153.00 per hour)</td>
<td>1.86</td>
</tr>
<tr>
<td>Maximum estimate (based on 12.14m hours @ €715.00 per hour)</td>
<td>8.67</td>
</tr>
<tr>
<td>Mean</td>
<td>5.26bn</td>
</tr>
</tbody>
</table>

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 report, 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
### 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:

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

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

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
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 100km\(^2\).

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)

\(^2\) 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:

<table>
<thead>
<tr>
<th>Title</th>
<th>Author (where known)</th>
<th>Publisher &amp; date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of the Orphan Works issue and the costs for Rights Clearance</td>
<td>Vuopala, Anna</td>
<td>European Commission, May 2010</td>
</tr>
<tr>
<td>Cost Models for Digitisation (PrestoSpace report)</td>
<td>Addis, Matthew</td>
<td>PrestoSpace, July 2005</td>
</tr>
<tr>
<td>Digitisation Costs &amp; Funding Workshop</td>
<td>N/A</td>
<td>OCLC, Oct 2003</td>
</tr>
<tr>
<td>eIFL Case Studies on Low Cost Digitisation Projects</td>
<td>de Vries, Repke</td>
<td>eIFL, Oct 2009</td>
</tr>
<tr>
<td>IFLA Conference Proceedings on Digitisation</td>
<td>N/A</td>
<td>IFLA, Aug 2009</td>
</tr>
<tr>
<td>INSAR Newsletter (issue 10)</td>
<td>N/A</td>
<td>Jun, 2005</td>
</tr>
<tr>
<td>JISC Digitisation Project Case Studies (various)</td>
<td>N/A</td>
<td>JISC, 2008-2009</td>
</tr>
<tr>
<td>NUMERIC Report</td>
<td>N/A</td>
<td>IPF, 2009</td>
</tr>
<tr>
<td>NUMERIC Report Appendices</td>
<td>N/A</td>
<td>IPF, 2009</td>
</tr>
<tr>
<td>The Electronic Archive Project in Hungary</td>
<td>Szatucsek, Zoltan</td>
<td>MOL, April 2009</td>
</tr>
<tr>
<td>The National Archives Commercial Digitisation Programme</td>
<td>N/A</td>
<td>SCA, 2009</td>
</tr>
<tr>
<td>The Status of Digitisation in Europe</td>
<td>Ayris, Paul</td>
<td>LIBER/EBLIDA, Feb 2010</td>
</tr>
<tr>
<td>Understanding the Cost of Digitisation</td>
<td>Hammond, Max &amp; Davies, Claire</td>
<td>Curtis &amp; Cartwright, Oct 2009</td>
</tr>
<tr>
<td>LIFE: Costing the Digital Preservation Lifecycle</td>
<td>Wheatley, Paul &amp; Ayris, Brian</td>
<td>British Library</td>
</tr>
<tr>
<td>Tracking the Reel World, a survey of Audiovisual Collections in Europe</td>
<td>Klijn, Edwin &amp; de Lusenet, Yola</td>
<td>European Commission, 2008</td>
</tr>
</tbody>
</table>

14.1.2 In the course of this study, we have also made reference to the following online sites/services (starts overleaf: ...)
<table>
<thead>
<tr>
<th>Organization</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akademie der bildenden Künste Wien</td>
<td><a href="http://www.akbild.ac.at/kuka/">http://www.akbild.ac.at/kuka/</a></td>
</tr>
<tr>
<td>Association of Moving Image Archivists</td>
<td><a href="http://www.amianet.org/">http://www.amianet.org/</a></td>
</tr>
<tr>
<td>Bibliotheque Nationale de France</td>
<td><a href="http://www.bnf.fr">http://www.bnf.fr</a></td>
</tr>
<tr>
<td>Centre National de l’Audiovisuel (CNA)</td>
<td><a href="http://www.cna.public.lu/">http://www.cna.public.lu/</a></td>
</tr>
<tr>
<td>Centre Virtuel de la Connaissance sur l’Europe (CVCE)</td>
<td><a href="http://www.ena.lu">http://www.ena.lu</a></td>
</tr>
<tr>
<td>Council of Europe</td>
<td><a href="http://consilium.europa.eu">http://consilium.europa.eu</a></td>
</tr>
<tr>
<td>Culturenet Denmark</td>
<td><a href="http://www.kulturnet.dk/en/text/">http://www.kulturnet.dk/en/text/</a></td>
</tr>
<tr>
<td>CultuurDatabank Vlaanderen</td>
<td><a href="http://www.cultuurnet.be/cultuurdatabank">http://www.cultuurnet.be/cultuurdatabank</a></td>
</tr>
<tr>
<td>Danish Film Institute</td>
<td><a href="http://www.dfi.dk">http://www.dfi.dk</a></td>
</tr>
<tr>
<td>Open Content Alliance</td>
<td><a href="http://www.opencontentalliance.org">http://www.opencontentalliance.org</a></td>
</tr>
<tr>
<td>ETHOS Toolkit</td>
<td><a href="http://ethostoolkit.cranfield.ac.uk">http://ethostoolkit.cranfield.ac.uk</a></td>
</tr>
<tr>
<td>European Film Gateway</td>
<td><a href="http://www.europeanfilmgateway.eu/">http://www.europeanfilmgateway.eu/</a></td>
</tr>
<tr>
<td>EuropeanaConnect</td>
<td><a href="http://www.europeanaconnect.eu/">http://www.europeanaconnect.eu/</a></td>
</tr>
<tr>
<td>EUScreen Project</td>
<td><a href="http://www.euscreen.eu/">http://www.euscreen.eu/</a></td>
</tr>
<tr>
<td>Finnish Museum of Photography</td>
<td><a href="http://www.fmp.fi">http://www.fmp.fi</a></td>
</tr>
<tr>
<td>International Association of Sound and Audiovisual Archives</td>
<td><a href="http://www.iasa-web.org/">http://www.iasa-web.org/</a></td>
</tr>
<tr>
<td>International Federation of TV Archives</td>
<td><a href="http://www.fiatfta.org">http://www.fiatfta.org</a></td>
</tr>
<tr>
<td>Joanneum Research (Graz)</td>
<td><a href="http://iis.joanneum.ac.at/iis/Default.asp">http://iis.joanneum.ac.at/iis/Default.asp</a></td>
</tr>
<tr>
<td>Kunsthistorisches Museum</td>
<td><a href="http://www.khm.at/">http://www.khm.at/</a></td>
</tr>
<tr>
<td>LIFE Project Website</td>
<td><a href="http://www.life.ac.uk">http://www.life.ac.uk</a></td>
</tr>
<tr>
<td>Maerlant Center</td>
<td><a href="http://www.maerlant.be">http://www.maerlant.be</a></td>
</tr>
<tr>
<td>National Cultural Heritage Agency</td>
<td><a href="http://www.kuas.dk">http://www.kuas.dk</a></td>
</tr>
<tr>
<td>Netherlands Institute for Image and Sound</td>
<td><a href="http://www.nibg.nl">http://www.nibg.nl</a></td>
</tr>
<tr>
<td>NUMERIC WS</td>
<td><a href="http://www.numeric.ws/">http://www.numeric.ws/</a></td>
</tr>
<tr>
<td>PrestoPRIME</td>
<td><a href="http://www.prestoprime.org">http://www.prestoprime.org</a></td>
</tr>
<tr>
<td>PrestoSpace</td>
<td><a href="http://prestospace.org">http://prestospace.org</a></td>
</tr>
<tr>
<td>UNESCO Audiovisual Archives</td>
<td><a href="http://www.unesco.org">http://www.unesco.org</a></td>
</tr>
<tr>
<td>UNESCO Communication and Information</td>
<td><a href="http://www.unesco.org">http://www.unesco.org</a></td>
</tr>
<tr>
<td>VideoActive</td>
<td><a href="http://www.videoactive.eu/VideoActive/Home.do">http://www.videoactive.eu/VideoActive/Home.do</a></td>
</tr>
</tbody>
</table>
European Commission

The New Renaissance
Report of the ‘Comité des Sages’ on bringing Europe’s cultural heritage online

Luxembourg: Publications Office of the European Union

2011 — 191 pp. — 14.8 x 21 cm

doi:10.2759/45571
How to obtain EU publications

Free publications:
• via EU Bookshop (http://bookshop.europa.eu);
• at the European Union’s representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu) or by sending a fax to +352 2929-42758.

Priced publications:
• via EU Bookshop (http://bookshop.europa.eu).

Priced subscriptions (e.g. annual series of the *Official Journal of the European Union* and reports of cases before the Court of Justice of the European Union):
• via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm).