

COMMISSION STAFF WORKING DOCUMENT

Accompanying the
**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS
on Europeana - next steps**

Part I: Questions for the public consultation

Part II: Overview of the current content contribution to Europeana by country

PART I

The present document complements the Commission Communication on 'Europeana – next steps' with a series of questions for consultation. Interested parties are invited to submit their comments on all or some of the questions **by 15 November 2009** to:

European Commission,
Directorate General for Information Society and Media
Unit 'access to information', EUFO 2281
Rue Alcide de Gasperi
L-2920 Luxembourg

or by e-mail to: ec-digital-libraries@ec.europa.eu

Contributions will be published on the digital libraries website of the Commission, unless requested otherwise by the organisation/person submitting them. The questions of this consultation all have a direct impact on the future development of Europeana. Some of the questions - in particular questions 7-10 - also have more general policy implications for the digitisation, accessibility and use of content from cultural institutions. The replies to these questions will feed into the further policy development in areas such as the re-use of public sector information and copyright related questions for the digitisation and online accessibility of cultural material. In the latter area the questions complement the work undertaken by the Commission in the context of the Green Paper on copyright in the knowledge economy and the follow-up Communication. Where relevant, the results of the consultation will be taken up and further discussed in stakeholder groups and working groups with Member States' representatives.

Questions for consultation

General

Question 1

Which orientations would you suggest for the future development of Europeana as a common access point to Europe's cultural heritage in the digital environment?

In our opinion, Europeana should seek the following goals mega-digital library with the following two orientations:

- 1- Become a respectable research library. This means Europeana will need to have on staff librarians that will be in charge of collections development, library services, etc.
- 2- Aggregate the diverse European heritage into a single access point with the active purpose to emerging as a cultural bridge not only between the European countries but to the rest of the world as well.

Question 2

Which features should be given priority in the further development of the site?

With new functionality and services being the focus of future upgrades, the quality of the digitized content on the site becomes vital. Search features and interfaces, and multilingual features bring the digitization workflow to the forefront. Creating digital assets that survive the passage of time and that can be programmatically repurposed to meet current and future access needs will ensure the value of the site and its content.

We believe a conservative list of key features that would make the Europeana library truly live up to its goal of becoming a useful and powerful digital library is as follows:

- Full text search inside the documents (most books are not OCR'd and hence this functionality is simply not possible to create unless these books are all re-processed through OCR)
- Persistent links to digitized material (a cursory testing of current site shows lots of broken links to books and images¹)
- Display of Table of Contents, Table of Illustrations, Index, etc to allow the user a more efficient access to the desired information.
- Navigation of books by chapter, and illustrations through direct metadata links. Only primitive pagination seems to be currently available.
- Homogeneous "look and feel" with a standard and consistent display interface, file format, functionality and image quality².

¹ Here's an example of a broken (amongst many) link:

<http://www.europeana.eu/portal/full-doc.html?query=descartes&tab=&start=1&startPage=1&uri=http://www.europeana.eu/resolve/record/01001/214FD824B4E1CE0A7E537DE63B4E9531B758146A&view=table&pageId=bd>

- Search inside the book field after a given book is accessed (again, requires that the book be OCR'd)
- Ability to extract one or more chapters from each book.
- Ability to extract particular specific illustrations from a given book.
- Ability to “copy and paste” directly from the page image (not the text mode), with bibliographic citation included.
- Ability to do full text search for books with relevance ranking

Unfortunately, most of the above features are not possible through site improvement and user interface upgrades, and can only be possible with a better quality of the digitized assets themselves. In other words, unless the digitized assets include high-quality images in standard formats, accurate OCR in XML (or other programmatically accessible format), and thorough structural metadata, no amount of site improvement can enable the above key features.

There are three essential elements in digitization that ensure longevity, multi-level access, interoperability and repurposing of the digitized assets:

- **Image quality** should include sharp text, color fidelity, high-resolution digital images in color, grayscale or black and white (bitonal) that are a faithful representation of the original. To ensure preservation of these images, their format should be in either TIFF or JPEG2000.
- Accurate **OCR** is the key to making text searchable. Also, unless this OCR is output into an open and tagged format, such as XML, the extraction, repurposing and interoperability of the OCR'd information could be compromised.
- Descriptive, technical and structural **metadata** are vital for tracking how images were created. Structural metadata, such as TOC, chapter headings, illustrations, index, etc. enables a level of navigation, rapid access, extraction and repurposing that the OCR alone cannot. . All the metadata should be included with the container of the digital images so they are integrated, thus the metadata can be programmatically extracted by the data base system manager.

Question 3

² The bulk of the Europeana books seem to come from the French Gallica website. Those books have been digitized with old manual scanning technologies and tend to be extremely poor in quality. Most have not been OCR'd, and even if they were OCR'd today, the results would be quite poor because of the image quality. See following example (amongst many) that illustrates the point:

<http://www.europeana.eu/portal/full-doc.html?query=balzac&qf=LANGUAGE:fr&tab=&start=2&startPage=1&uri=http://www.europeana.eu/resolve/record/03504/9117450F9B14373051420D94391E6C981AF7AD46&view=table&pageId=bd>

Has Europeana struck the right balance between making Europe's digitised cultural heritage searchable through a common entry point and at the same time giving visibility to the institutions that contribute the material, or should the material accessible through Europeana be presented in a more unified way?

We found the hopping from site to site confusing and presents too much variability while not allowing the customary navigation of the Web such as the ability to go back and forth within the same browser window. We believe that the user's experience should take a clear precedent over the "bragging rights" of the member institutions, and that can only be achieved, in our view, through a single, unified and self-contained access point. It is worthwhile to consider having the language for the site be a common language and also have the native language of the country from which an object came from. Furthermore, the images can be watermarked with the country of origin.

Question 4

How should Europeana further develop its own autonomous identity?

In order to truly evolve into a European Digital Library, Europeana needs to realize that it takes a lot more than a large number of scanned material to achieve that status.

Europeana should:

- have its own Head Librarian on staff,
- be “patron-centric” in its development roadmap
- address “Collection Development” based on gaps and needs, and not purely on opportunistic initiatives
- Develop and maintain a consistent “brand” in terms of user experience, “look and feel”, and quality of product and services

Question 5

Should there be minimum requirements for the content brought into Europeana by the contributing organisations (e.g. minimum viewing or use options)? If so, who should be responsible for defining and imposing these minimum requirements?

Absolutely! As indicated in Question 1, the quality of the digitized assets has a direct impact on the viewing and use options that can be made available. Whether Europeana chooses to develop its own unified viewing interface or leaves it to the member institutions, the user experience must dictate at least a minimum set of requirements that the digitized assets must meet in order to enable a baseline user experience. So in other words, the minimum requirements should be defined by the following criteria:

- **Longevity** of the digitized assets to ensure that today’s investment will not become obsolete in the near future. this means quality that is industry standard or better, and formats that are industry standards
- **Access** in multiple forms to meet the current needs of the user base. access for view and access for quality printing may required two versions of the images for good reasons and patron satisfaction.
- **Repurposing** to enable new forms of access that are either emerging or will emerge in the future, e.g ebooks in epub format. Extensive metadata now is insurance to meet these unknown needs of the future.
- **Inter-operability** to ensure that digitized assets from different member institutions can be combined together in a consistent and unified way

The above four high-level criteria will then dictate the lower level specifications for the contributed content

Content for European

Question 6

Which categories of content are so important for the users that Member States and their cultural institutions should be encouraged to make them available through Europeana?

What measures can be taken to ensure the availability of these works through Europeana?

All categories without any exception should be digitized, and starting with special collections that have not been digitized yet, and that present the least redundancy among the member states should be a priority. Because the market place is global, the desired categories, from the user perspective, most likely span the full range of categories. Also European libraries and archives hold in their libraries a great amount of invaluable and unique material about their former colonies. Hence that content is of great value to many countries outside the European continents, and should also be a priority.

Question 7

What is the best way to encourage cultural institutions and rightholders to take into account cross-border access - including through Europeana - in their agreements on digitisation and dissemination of in-copyright material? Which legal or practical barriers to this cross-border access need to be addressed?

Well thought-out and carefully written legislation for copyright protection of the holder balanced with serving the public access goal for materials about the culture and heritage of the country. Perhaps nominal viewing fees not unlike internet “click” charges. Researchers, students, others could have a pre-paid account to allow access and view privileges that could be shared with the copyright owners.

Question 8

How can the difference in the level playing field for digitising and making accessible older works between the US and Europe (in particular the 1923 cut-off date in the US, that places all material from before 1923 in the public domain) be addressed in a pragmatic way (e.g. better databases of orphan and out-of-print works, a cut-off point that imposes lower requirements for diligent search in relation to orphan works)?

Currently it costs 4 to 5 times more to conduct due diligence on a single orphan title than to digitize that title. Hence it is not an affordable or scalable process. It's also very impractical to use the year of the death of the author as the base for the cut-off date. Again, we believe new legislation is needed for both changing the cut-off date calculation as well as allowing the digitization (and hence preservation) of all the orphan works with a fair remuneration process and opt-out options to the potential right holders.

Question 9

What policies should be adopted to avoid that the process of digitisation itself creates new types of sui generis copyright that, in turn, could create barriers to the dissemination of digitised public domain material?

In our opinion, a “first mover” advantage is key to discouraging unwarranted attempts or de-facto copyrighting of public domain material. As long as the public domain material is of high quality and widely and freely available from non-profit digital libraries, then we believe this will eliminate any risk of that content re-entering the copyright domain again. There may however be incremental value for the end user to have access to that content in other derivative forms, such as Print-on-demand, ebooks, and other outputs that require additional labor and costs and hence can only be offered for a fee that covers those costs and allows for a reasonable profit. We believe those are positive developments that will further help disseminate that content, not create barriers to it. This would be reminiscent of the open source market, where open source software such as Linux is freely available, but value-added integrators like Red Hat sell customized versions thereof along with support services to Linux users. Also, the Open Source movement has proven that self-imposed policies and rules of conduct can be effective in preventing open creative works from becoming proprietary. One could certainly envision a similar set of self-imposed regulations and policies for the public domain content.

Question 10

*What measures can be taken to ensure that cultural institutions make their **digitised public domain material** accessible and usable in the widest possible way on the Internet? Should there be minimum requirements for the way in which digitised public domain content is made available through Europeana?*

Unfortunately, there exists no common digitization standards that are endorsed and followed by the key players of digitization around the world. Europeana could be the one golden opportunity for the whole world to finally help enforce digitization standards that ensure (i) the longevity, (ii) inter-operability, (iii) multiple-access formats and (iv) long-term repurposing of that digitized content. Absent any one of these criteria, we fear that the value of that content will quickly decline over time and become obsolete.

Financing and governance

Question 11

Which financing model would reflect a fair distribution between Community funding, Member States' funding and private funding, and taking into account that the aim of Europeana is to give the widest possible access to Europe's cultural heritage at pan-European level? Could Europeana be financed solely by national cultural institutions or by private funding?

If Europeana is to stand as a mega Digital Library that serves Europe's cultural heritage to an access level that no other physical library could ever aspire to do, is it not then worthy of at least the same level of investment, attention and care as any National Library of some of the largest Member States? If so, then why can't the EU allocate a budget similar in size and scope to that of building and staffing a library of similar scale, except that in this case it would be a digital library. The cost of that project is almost certain to be lower than any of the national libraries of France, England, Germany or Italy.

As for private funding, while it may be a source of greatly needed funds to accelerate the growth of Europeana, it must come with no strings attached, and in no way should it threaten the quality of the digitization, the independence or the mission of the Europeana program.

Question 12

Is sustained European Union funding for the basic operations of Europeana necessary and justified for the period after 2013? What type of European funding instrument could best be used?

In our humble opinion, and given the state of Europeana today, it's unrealistic to expect the program to be near completion by 2013, and hence funding should be sought, increased and will be necessary long after 2013. Europeana must also start thinking about not only the development cost, but also maintenance, sustainability, and updates costs. This is not unlike the planning and thinking about the costs for sustaining a physical building that has high volume public traffic. When presented with the daunting task before Europeana, digitization of all content, and the maintenance and expansion thereof is expected to require well beyond 2013.

Question 13

Which governance structure for Europeana would best fit the preferred financing model (as indicated under question 11)? Should there be a role in the governance structure for organisations other than content providers?

No comments.

Question 14

How can private involvement in Europeana best take shape (e.g. through sponsoring, through technological partnerships, through links from Europeana to the sites of publishers and other rightholders where the user can buy in-copyright content, or through another type of partnership)?

All of the above would be welcome vested commitments that would enhance the value and dissemination of that digital content. In-copyright content would almost certainly have to involve a reasonable profit model that is attractive enough to publishers and right holders, but still delivers unique value at a reasonable price to consumers and institutional customers. It might also allow to resolve the thorny, but nonetheless urgent, copyright concerns around the so-called Orphan Works that risk of deteriorating beyond recovery if no action is taken to save them.

Question 15

How can private sponsorship of Europeana best be stimulated? Are commercial communications on the Europeana site acceptable, and, if so, what type of commercial communications (e.g. logos of sponsors, promotion of specific products)?

Our humble opinion is that what matters more is less than the form but rather the substance of the commercial communications and their relevance to the mission of the Europeana program. If those communications, albeit of commercial nature, bring real value, enhance the mission of the site and serve the needs of the consumers, then those would be desirable. Otherwise, they might undermine the value and credibility of the site, and hence would be undesirable.

Question 16

Should there be a contribution (financial or other) in exchange for the links from Europeana to sites with content for which the user has to pay? Can a model such as that of Gallica 2, providing links from the site of the Bibliothèque Nationale de France to the content on the sites of French publishers, be transposed to Europeana?

If Europeana is going to serve as a conduit of users to paid content, then it's only fair and highly advisable for it to receive an appropriate contribution that could be used to further invest and enhance its program.

PART II

Overview of the contribution of Member States to Europeana in % of the total number of objects (situation end July 2009).

EU countries

France	47%
Germany	15.4%
Netherlands	8%
United Kingdom	7.9%
Sweden	5.2%
Finland	4%
Greece	1.6%
Italy	1.2%
Belgium	1.1%
Slovenia	0.7%
Spain	0.6%
Estonia	0.4%
Luxembourg	0.4%
Poland	0.3%
Romania	0.3%
Austria	0.2%

Portugal	0.2%
Hungary	0.1%
Latvia	<0.1%
Cyprus	<0.1%
Bulgaria	<0.1%
Czech Republic	<0.1%
Denmark	<0.1%
Ireland	<0.1%
Lithuania	<0.1%
Malta	<0.1%
Slovakia	<0.1%
Non EU countries	
Norway	4.3%
Switzerland	0.4%