



Towards new Data Provider and Aggregator Agreements

Europeana Foundation implemented Data Agreements at the end of 2009 to reach common agreement on the use of the metadata of its content providers and aggregators, and be able to become an operational service. 85% of providers have signed these agreements giving Europeana the right to reuse the content in non-commercial circumstances. These agreements were necessarily a compromise, to reach consensus and move forward, but will not serve Europeana, or other aggregators, for the future.

Linked Open Data is needed in Europeana for almost everything it proposes to do from Danube onwards. To create the semantic connections so that different representations of the same thing can be found, to ensure that the trusted, authentic cultural objects can be found on the sites where they belong, to promote the value of cultural heritage to the 21st century generations.

To allow the publishing of the amalgamated metadata of Europeana as Linked Open Data requires the non commercial clause, of the current agreements, be removed. When published as Linked Open Data, the metadata could be reused for commercial purposes and Europeana has no means of controlling this.

There is an opportunity for the Cultural and Scientific Heritage Institutions to make an enormous difference to the quality and trustworthiness of the Linked Open Data layer. Rather than leaving the space open to the less well qualified creators of such data, Europeana can by releasing its data as open and interoperable, promote the importance of the contributing institutions in the knowledge economy. Linked Open Data will improve the discovery of material held in the cultural heritage institutions. . Acknowledging this, institutions such as the British Library has recently taken the bold step of publishing its Bibliography as CC0.

Additionally, politically, Europeana needs to remain a flagship while at the same time, developing widespread use of the metadata it amalgamates. To do both of these Europeana needs to work with the web as it is now and as it is shaping up. This requires the placement of data in an open context on the web, the use of search engines to market data to the end user and the development of partnerships that give the data context and excite the end user. The simplest way of achieving this is to make her data standardized and open and accessible to those who need it, to create a thriving knowledge economy in cultural and scientific heritage.

Creating the new agreements

A series of workshops, with representatives, from the network were held to produce these new version of the agreements.. These were organised by EuropeanaConnect WP4, the Europeana Licensing Framework,.. The workshops reviewed the limitations of the current Data Provider & Aggregator Agreements in the context of Linked Open Data, use of Wikipedia and the search engines as promoters of Europeana and the Content Provider & Aggregator sites.

The first workshop was held on the 30th September 2010 and addressed those providers who had found signing the current agreement most difficult. This was followed by 4 domain workshops, Museums, Archives, Libraries and Audiovisual. For each workshop a poll was taken at the beginning to find out if the delegates thought the non commercial clause in the agreements could be removed.



The results were often 50:50 or the majority in favour of keeping the non commercial clause. The workshop would then work through the risk and rewards of open data and end up with a poll at the **end showing that the majority felt that the non commercial clause needed to be removed** in the interests of being able to publish metadata as Linked Open Data. The workshop participants, as representatives of their communities, also worked iteratively on the Data Provider Agreement itself. This led to the new draft version that is sent out today for wide consultation. This version removes the non commercial clause. This draft version is sent today to the 200 strong Thematic Network of Europeana. A period of 3 weeks is being allowed for initial consultation with comments due back by the 7th of January. On the 17th of January a workshop will be held with the 'Aggregators' to talk about the results of the wider consultation and what would need to be done to persuade their providers to accept such an agreement. A second round of consultation on the resulting Agreements will be undertaken during February. It is hoped to agree on a new set of Data and Aggregator Agreements by the beginning of March to be able to publish a first set of Europeana Linked Open Data.

Risks and rewards

Notwithstanding the technical difficulties publishing Linked Open Data might involve, the main issue was about a balance of the risks and rewards. Below is a summary of the risks and rewards that participants raised at the workshops.

Risks	Rewards
About the institution's position	
Loss of control and authority	Greater relevance for the institution in the digital era. By opening up access to my data I avoid creating yet another online silo.
Loss of branding	Possibility to share the workload and gain more support from the community
Loss of the connection with the institution	Increased accessibility to my online content
Loss of control over the channels of access and the KPIs to justify use	Greater fulfillment of the mandate to share knowledge with the public
Fear of damage to the brand and reputation of my institution if my data is placed in defamatory or illegal contexts.	This fear exists anyway with putting anything online.
Fear of the new	Reputation and publicity in being the first. Other institutions ask for my expertise.
Makes users more lazy. They think that if it's not on the web it doesn't exist. People use material without the appropriate training.	Driver for digitisation and more sophisticated searching techniques.
About data quality	
Loss of control over the integrity of the data	Exposure puts more pressure on quality
Loss of control over the presentation and display	There are more, often unexpected functionalities and services that spring from the use of their data which add usability
Quality decrease	Quality increase
No real-time updates	
Loss of context	Improved context. Data enrichment. Building a web of data. I discover more things about my content than I knew or was able to do before.
We hold content that doesn't belong to us. We	Not everything needs to be given away. Only



hold content that we don't know the provenance/the status of	what an institution feels comfortable with or is capable of making available without compromising 3 rd parties' contributions.
Financial risks	
Enable a commercial activity with something created with tax-payers' money	What is made available with tax-payers' money should be made widely available to all. Besides, almost everyone accepts to be crawled by Google-the biggest online commercial activity of them all.
Copyright enables money-making which is important at a time of important financial cuts	More money is made because of increased exposure and traffic. It is about making the pie bigger and not about cutting it down in smaller slices...
	More real visitors to an institution's physical space.
	Giving something away for free doesn't mean that something can't be commercialised elsewhere or in some other way. Someone that wants to use a picture in a print edition will always be going back to the institution holding the rights to license the right to use a high-res image, for example.
	Increased opportunities for funding. A lot of funding bodies are nowadays supporting open access.
	Saving costs from sharing workload. I can complement my collection with information I get from other bodies that are also making available their resources.
Technical	
More work and infrastructure required by my organisations	Europeana will do most of the work for you. We hope that more traffic is not seen as a risk!

The question for each organisation to answer is whether they think that the risks outweigh the rewards.

At the workshops there was a general feeling that risks most often involved perceived fears while the rewards involved more tangible real benefits. Participants also acknowledged that there is a general trend towards openness that data providers need to observe and there is some urgency to be on top of developments in order to remain relevant in the digital era.

Questions raised at the workshops

Please find below some of the Frequently Asked Questions of the workshops with some explanations and answers, to help you in your understanding of the need for change. .

Why is Europeana so keen on changing the current agreements?



The current agreements are very restrictive for Europeana. For instance, it is not possible to place the data on Wikipedia or publish them as LOD. In principle, even to allow for the API use we have to assess every time if a target website is commercial or not. In addition:

- by adopting a standard license for the metadata through our new agreements we can achieve greater data compatibility on the web.
- traffic to partners' sites will be enhanced by providing structured metadata to Google.
- publishing the data as LOD through one single point such as Europeana enhances the collective potential of our resources.

Why is it so important to remove the Non Commercial clause?

Because it is stopping Europeana and other aggregators ability to compete in real world.

Current agreements say: " the Data Provider grants Europeana Foundation the right to allow Third Parties for non-commercial purposes only and on a non-exclusive, worldwide, royalty-free basis to...". Trying to identify what is a non-commercial purpose would be a headache. There is probably no purely non-commercial activity, or as one workshop participant put it, 'there is always a coffee corner in every public library'

So it would be difficult to enforce. It is also a fact, that almost everyone is crawled by Google these days to increase the traffic to their websites and Google is a purely commercial entity that displays ads next to your content. The workshops also revealed that partners are afraid about some imagined loss of income or about someone else commercializing on data created with tax payers' money and yet have a public remit to provide information for the public good. This cannot be solved by the need to remove the non Commercial clause, but its removal does not prevent money being earned at a later stage from the data.

What is the idea behind Europeana publishing its data as Linked Open Data?

Europeana Foundation exists to give access to the content held in cultural heritage institutions across Europe. It will do so according to the wishes of the majority of those institutions and within the law. Therefore, Europeana will publish the metadata she collects as Linked Open Data only when consensus is reached with the majority of the providers. We hope that the current consultation will help raise awareness about the benefits of Linked Open Data and give us the buy-in to start publishing our metadata around March. The idea is not to make money out of changing the agreements to non commercial, just to be able to make the metadata work better as an advertiser of the content held by the providers.

What is happening to my metadata once it's published as Linked Open Data? Is it possible for someone to download the whole bulk of my data? Is it possible to download the whole bulk of Europeana's data?

After your data is published as Linked Open Data, it will be possible to enable the semantic potential of your resources and interlink them with other trusted resources within and outside Europeana from the growing Linked Open Data cloud¹. It is technically possible to download the whole of Europeana's and of therefore your metadata if it is published as Linked Open

¹ <http://linkeddata.org/>



Data. However the workshops of your colleagues, thought this to be very unlikely and that the risks of it happening are outweighed by the rewards.

I might be infringing copyright by making my data openly available

In principle metadata contains factual information and therefore, no copyright applies. If a piece of metadata includes significant knowledge, as in a curated description, this could be excluded from the metadata contributed to Europeana and thus from being made openly available. With regards to the sui generis Database Right, a provider can decide to waive it by adopting an appropriate open license. As for privacy, content providers are asked to observe the laws and regulations applicable in their own country.

What about the metadata we receive or buy from 3rd parties or the use of 3rd party resources in our metadata such as proprietary thesauri?

You should make available to Europeana only the metadata that you feel comfortable with giving and for which you have the rights cleared. If you are using 3rd party resources you should clear the rights or remove this information from your metadata.

Is Europeana going to make money with our metadata?

The Data Provider and Aggregator Agreements aim to give best practice for the reuse of others' metadata. The lifeblood of Europeana is the metadata provided by the content providers and therefore the relationship with content providers must be built on openness and trust. Europeana seeks to change the current agreements because they are hampering it from fulfilling its mission of increasing visibility and contextual relevance for the digital content of Europe's Cultural & Scientific Heritage institutions.

Europeana will therefore not make direct use of the metadata, for commercial gain, without prior consultation and agreement with the data providers themselves.

Participation in Europeana, by way of contributing metadata to advertise the content held in institutions, does not preclude the sale of that metadata to a commercial company by the content provider themselves. Indeed the workshop to create value propositions with commercial providers had them asking for premium, paid access to such material.

Will Europeana change its' contract again very soon?

We understand content providers' and aggregators' concerns and we hope that the current agreement meets the needs of Europeana and the wider network for the future. However, in a fast changing online environment, we cannot know for sure when new developments will raise the need for new agreements.

Why did Europeana remove the reference to the full text?

Europeana wanted to simplify the agreements and remove from the current formulations any references to content even if relevant clauses referred to use of content just for indexing purposes. No full text has been delivered so far and any such activity will be covered in the future by separate agreements.



What about the current agreements? What happens next?

We hope to have the new version of the agreements ready for signing by March. We would aim to convert everyone to the new agreements by July. In the meantime the old agreements will be valid.