

→ **The Problem of the Yellow Milkmaid**
A Business Model Perspective on Open Metadata

Authors

Harry Verwayen, Europeana, The Hague (NL)

Martijn Arnoldus, Kennisland | KnowledgeLand, Amsterdam (NL)

Peter B. Kaufman, Intelligent Television, New York (US)





‘The Milkmaid’, one of Johannes Vermeer’s most famous pieces, depicts a scene of a woman quietly pouring milk into a bowl. During a survey the Rijksmuseum discovered that there were over 10,000 copies of the image on the internet—mostly poor, yellowish reproductions¹. As a result of all of these low-quality copies on the web, according to the Rijksmuseum, “people simply didn’t believe the postcards in our museum shop were showing the original painting. This was the trigger for us to put high-resolution images of the original work with open metadata on the web ourselves. Opening up our data is our best defence against the ‘yellow Milkmaid’.”

1. Executive Summary

Interest in open metadata is growing among policy makers, the cultural heritage sector, the research community, and software and application developers. At the European level, the Digital Agenda for Europe 2020 identifies ‘opening up public data resources for re-use’ as a key action in support of the Digital Single Market.² The European Commission is reviewing the Directive on Re-Use of Public Sector Information. The Commission’s The New Renaissance report³, published in January 2011, emphatically endorsed open data. At the national level, for example in the UK, the higher education community has issued the Open Metadata Principles⁴ calling on metadata to be openly available for innovative re-use.

For the past 12 months Europeana has been exploring with its partners the issues surrounding open metadata, in the belief that openness brings benefits both to the cultural heritage sector and to the broader knowledge economy. This position is echoed by the Vice President of the Commission responsible for the Digital Agenda, Neelie Kroes, who has declared: ‘I urge cultural institutions to open up control of their data...there is a wonderful opportunity to show how cultural material can contribute to innovation, how it can become a driver of new developments. Museums, archives and libraries should not miss it’ (Kroes, Neelie 2011).

It is in this context that Europeana, together with its contributing partners, has spent the last year reviewing its Data Exchange Agreement, which governs the rights under which the metadata from Europe’s cultural heritage institutions is made available in its repository. One of the most important changes in this new agreement is that it calls for a more open licence (Creative Commons CC0), which allows for the re-use of descriptive metadata in a commercial context or by commercial players. This change of agreement is necessary for the development of Europeana, which has successfully proven the value of its supply-led business model in aggregating massive data sets from all domains across 32 countries. But to be able to achieve sustainable success in the crowded content arena of the Internet, Europeana must now move to a demand-led model, positioning itself as a distributor of data and facilitator of digital

¹ <http://bit.ly/mRoOfp>

² <http://bit.ly/k97K8c>

³ <http://bit.ly/tafh4T>

⁴ <http://discovery.ac.uk/businesscase/principles/>



heritage R&D in accordance with its Strategic Plan⁵.

Europeana's extensive consultation with the heritage sector, including dozens of workshops, has explored in detail the risks and rewards of open data from different perspectives. The most helpful way of framing this discussion has proven to be around the business model of cultural heritage organisations. The findings in this white paper are drawn from a July 2011 workshop in which key actors from museums, libraries and archives evaluated their metadata within the context of their own business model.⁶ Placing metadata within their business models gave workshop participants the opportunity to assess the monetary and reputational utility of metadata to their respective cultural organisations.

Participants in the July 2011 workshop in The Hague, The Netherlands

Roei Amit	INA, France
Martin Berendse	National Archive, The Netherlands
Caroline Brazier	British Library, UK
Mel Collier	Leuven University, Belgium
Jonathan Gray	Open Knowledge Foundation, UK
Renaldas Gudauskas	National Library of Lithuania, Lithuania
Lizzy Jongma	Rijksmuseum, The Netherlands
Peter B. Kaufman	Intelligent Television, USA
Caroline Kimbell	The National Archives, UK
Jan Muller	Sound and Vision, The Netherlands
Lars Svensson	German National Library, Germany
Helmut Trischler	Deutsches Museum, Germany
Bill Thompson	BBC, UK

The workshop participants differentiated between three types of business models for dealing with metadata. In most cases metadata is created as part of the **public mission** of the institution, and it has no direct or indirect effect on the value creation and revenue streams of that organisation. Quite often, however, metadata can be seen as a **key activity** of the organisation as it contributes indirectly (as a marketing tool, for example) to the revenues of the organisation. A few cultural heritage institutions derive revenues directly from the creation and selling of metadata; metadata then becomes then a core **value proposition** of the organisation.

Differentiating between the different roles of metadata in the business model helped frame the benefits and risks associated with open licences. Opening up metadata under open licensing terms will have different effects, depending on the roles that metadata plays in these business models. In the first two cases, opening up metadata was seen to have largely positive effects (more widespread use and visibility of the content) and limited negative ones. It is when organisations earn money directly from selling metadata that there is a potential for negative effects.

A critical factor in these cost-benefit evaluations is time. Workshop participants broadly agreed that 'over time, the benefits will no doubt outweigh the costs'. Participants also recognised the serious risk of memory organisations being sidelined as application innovations gather momentum and developers focus mainly on openly

⁵ <http://bit.ly/fCunRH>

⁶ Osterwalder, A., & Pigneur, Y. (2009) *Business Model Generation*, New Jersey



licensed datasets. If cultural heritage organisations do not expose data in ways that digital natives want to use it, they risk becoming irrelevant to the next generation.

For organisations opening up their metadata, three major advantages were identified. These were the **increase in their relevance** to digital society, the **fulfilment of their public mission** to open up access to our collective heritage, and finally the value of opening up **access to new users**, who are prompted to engage with the object in its digital form and subsequently with its real-world source.

We recommend that three specific issues need to be addressed:

1: Loss of Revenue/Spill-over Effects: Opening up data should be seen as an important part of the responsibility of our public cultural sector. Instead of measuring success by the amount of commercial revenue that institutions are able to secure from the market, new metrics should be developed that measure the amount of business generated (spill-over) based on data made openly available to the creative industries. This requires a change in evaluation metrics on a policy level.

2: Loss of Attribution: Heritage institutions are the gatekeepers of the quality of our collective memory, and therefore a strong connection between a cultural object and its source is felt to be desirable. There is a fear that opening up metadata will result in a loss of attribution to the memory institution, which in turn will dilute the value of the object. Investigations need to be made on the technical, legal and user levels to safeguard the integrity of this data.

3. Loss of Potential Income: A very limited number of institutions currently earn significant money selling metadata. It has been argued that the loss of this income can be averted by product differentiation: data can be made available openly in one format and marketed in another format under commercial terms. A larger issue is the fear of losing the ability to sell data in the future when data is openly available for everyone to use. This requires a change of mindset, acknowledging that, in reality, we are all invited to create new, commercial services based on open data.

Overall, the **conclusion** of the workshop participants was that the benefits of open sharing and open distribution would outweigh the risks. In most cases the advantages of increased visibility and relevance will be reaped in the short term. In other cases, for example where there is a risk of loss of income, the advantages will come in the longer run and short-term fixes will have to be found. All of this requires a collective change of mindset, courage to take some necessary risks and a strong commitment to the mandate of the cultural heritage sector, which is to enable society to realise the full value of the cultural legacy that is held in the public realm.

2. Introduction

Europeana is currently revising the 2009 Data Exchange Agreement that governs the way its 1,500 partners – museums, libraries, archives, holding millions of images,



texts, recorded sounds and moving images – license their metadata to Europeana.⁷ Europeana uses that metadata – ingests it, indexes it, enriches it, makes it available online – in order to aggregate and expose the associated digital cultural content its data providers collect, curate, and host.

To marry the long-term societal and macro-economic benefits with the short-term interests of institutions requires a different way of thinking about the roles, responsibilities and business models of cultural heritage in the digital age. For many of our memory institutions the creation of metadata has been an important part of their organisational activities for centuries. Opening up access to our collective memory through the release of that data can be argued to be an intrinsic part of their responsibilities. While most institutions agree to this in principle, the day-to-day realities in which they are operating make them hesitant to bite the bullet when it comes to opening up their data. This reluctance stems in part from the pressure from policy makers on cultural institutions to generate revenue from the market, a lack of clarity about the legal situation in which they are operating, but mostly from a fundamental lack of trust that the benefits will outweigh the costs associated with the innovation of their business model.

This is all quite understandable, especially in this volatile and insecure economic climate. Nevertheless, participants in the July 2011 workshop argued that we need to face these challenges head on and find solutions urgently, as the alternative is altogether more unattractive.

“We risk consigning ourselves to irrelevance in 15 years time.”

This business/economic basis for promoting open access to cultural and educational information is in turn situated in the context of what analysts, journalists, and academics have described now for 10 years as the new “political economy of open source.”⁸ These analyses are not rooted in the so-called copyleft movement, which often features arguments that disparage intellectual property as a concept and decry patents and copyrights as useless encumbrances upon human progress in the digital age. Rather, these are business arguments put forward by many of the leading technology companies in the world – IBM, Oracle, Nokia, Cisco, Microsoft – who open their development environment to broad communities that can enhance, and now are vigorously enhancing, the value of their products and services.⁹ Although it is difficult to draw many direct connections between building commercial value in the software and technology business, on the one hand, and work in the cultural sector by museums and libraries, on the other, both do have much in common when it comes to the value they can build for their own enterprises and institutions by paying attention to what can be described as the economics of innovation,¹⁰ and of user innovation in

⁷ The data-exchange agreement states that “whatever data is given to Europeana is called...metadata.”
<http://bit.ly/jox4iy>

⁸ Weber, Steven (2004), *The Success of Open Source*, Cambridge, Harvard University Press.
<http://bit.ly/u48nmc>

⁹ <http://www.economist.com/node/5015177> 20 May 2005

¹⁰ <http://www.rogerclarke.com/EC/Bled04.html>



particular¹¹. Intelligent Television and Creative Commons Netherlands are co-producing a television documentary on this subject for 2012.

“Over time, the positives will no doubt outweigh the negative effects of opening up data.”

The Business Model Canvas

In the context of this white paper a business model is understood to be ‘the rationale of how an organisation creates, delivers and captures value’ (Osterwalder, Pigneur 2009).

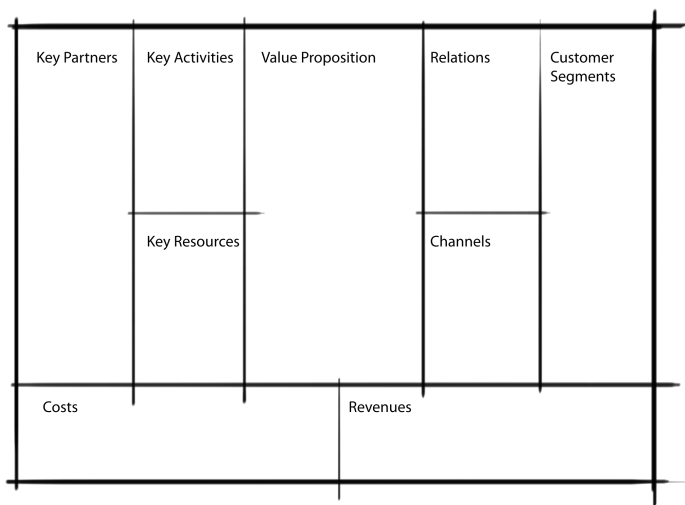


Figure 1: Business Model (Osterwalder, Pigneur 2009)

The theoretical framework of a business model consists of nine interrelated building blocks that depict the logic of how the organisation intends to deliver value:

1. Customer segments: an organisation serves one or several customer segments.
2. Value proposition: an organisation seeks to solve customer problems and satisfy customer needs with value propositions.
3. Channels: value propositions are delivered to customers through communication, distribution and sales channels.
4. Customer relationships: each value proposition offered to a client group establishes a relationship.
5. Key activities: the activities that are required to offer and deliver the value proposition.
6. Key resources: the resources that the organisation needs to perform the said activities.
7. Key partnerships: the partnership network the organisation needs to establish to perform certain activities that it cannot efficiently perform by itself.

¹¹ Von Hippel, Eric. (2005) *Democratizing Innovation* Cambridge, MIT Press.; free online at: <http://web.mit.edu/evhippel/www/books.htm>. Intelligent Television and Creative Commons Netherlands are co-producing a television documentary on this subject for 2012.



8. Benefits: the building blocks are organised in a front end (the 'what' and the 'who') that defines the revenue building capacity of the organisation.
9. Cost structure: the back end, or the 'how' of the business model, establishes the cost structure of the organisation.

Europeana currently holds metadata that links to cultural heritage on the sites of content providers. Although an ambiguous term¹², metadata in this context refers to the data that describes the key characteristics of the actual content; for instance the name of the work, its creator, date of creation and other background information. Metadata can either be quite plain or extensive and specialised. When users search the metadata that Europeana holds, the result links them back to the digital content on the original data provider's website. The cultural institutions that have provided data to Europeana to date have done so under the conditions specified in the original Europeana Data Agreement issued in 2009. One of the conditions governing this metadata is that it can be re-used only for non-commercial purposes.

3. Supply and Demand

From a business model perspective, the aggregation of this collection of metadata in the period 2008-2011 can be seen as an effort driven by the supply side resulting in a repository that currently holds over 20 million metadata records from over 1500 institutions. This is not a small feat from an organisational, technical and legal perspective.

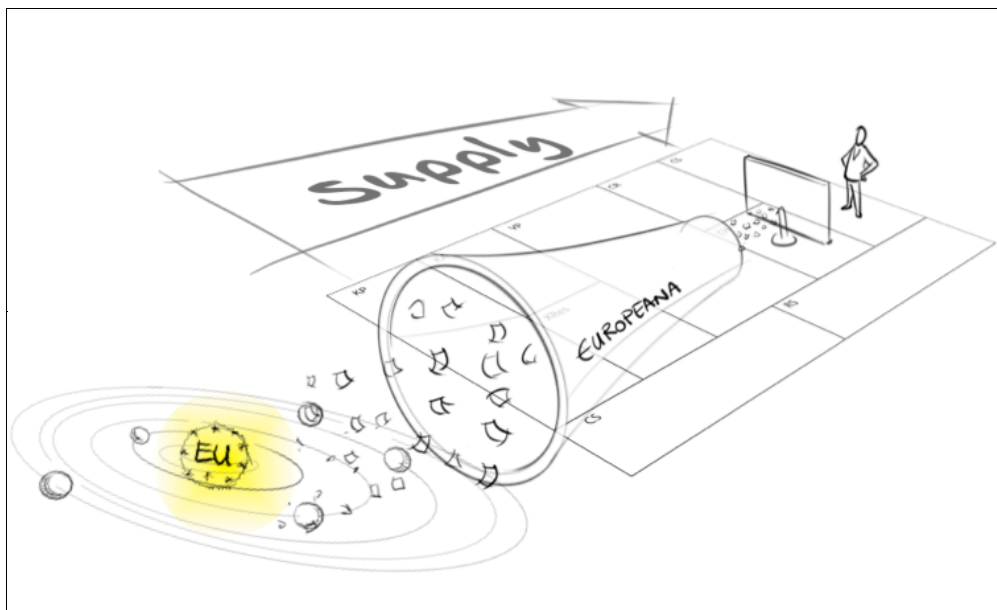


Figure 2: Supply-driven business model

During this 2008-2011 phase, much of the work of Europeana was focused on the 'back end' of the business model: setting up technical infrastructure for aggregation, standardising metadata formats, fostering a network of participants, etc. The data was made accessible primarily through the portal Europeana.eu, which complies with the

¹² <http://en.wikipedia.org/wiki/Metadata> 2 November 2011



legal framework of non-commercial use. With the infrastructure now largely in place, the business model focus of the organisation has to shift to a more user-oriented, demand-driven business model. This requires a very different set of skills, organisational setup, value propositions and legal framework: **the ability to re-use information found on Europeana consistently comes up as the number one priority for end users**¹³.

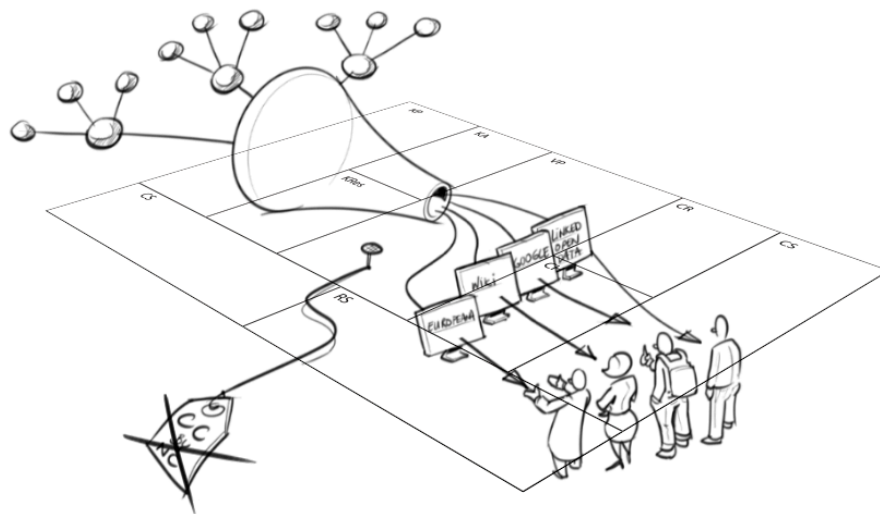


Figure 3: Demand- driven business model

4. Open Licences

This change in focus from supply-led to demand-led is brought about by two imperatives.

The first is the impetus to open up Public Sector Information so that it can generate new applications and services, thus furthering the Digital Agenda for Europe and fulfilling the recommendation in The New Renaissance report that the metadata related to digital objects produced by cultural institutions should be widely and freely available for re-use.¹⁴ Europeana is redrafting its Data Exchange Agreement in the context of other public service digital information that is being created with taxpayer money (most of the cultural institutions providing information to Europeana are similarly funded) and the European Commission's new guidelines on the re-use of that information, and the public benefit of that re-use. These guidelines, which have

¹³ IRN research (2011), *Europeana Online Visitor Survey*: <http://bit.ly/siAF8r>

¹⁴ *The New Renaissance*, ibid p.5



been adopted by all 27 member states of the European Union, stress how public information has economic potential to generate new businesses and jobs.¹⁵

The second is the requirement, in the crowded online content market, to provide material on the terms that users want it, in the places they need to use it. To do less is to fail to compete; to fail to compete is to limit the sustainability of Europeana as a flagship enterprise for the display of Europe's cultural diversity and treasury.

Europeana began, of course, as an effort to create an online European library that would make Europe's cultural heritage accessible for all.¹⁶ But in the years since that initial vision – years that include the launch of the European digital library prototype in November 2008; the launch of Europeana version 1.0 in February 2009; and the current version of the site, providing information on more than 20 million digital objects – the web and the technology of digital content have evolved with accelerating velocity.

Today Europeana acknowledges the importance of linked open data for its future: of allowing – facilitating – the creation of semantic connections through the harvesting of data, application programming interfaces (APIs), and other technological innovations to help users access information from authoritative sources about cultural objects. However, the original Data Exchange Agreement grants Europeana use rights only for non-commercial purposes. Linking providers' data across websites that carry advertisements (Google Adwords, for example, or traditional display ads) or across applications developed for commercial social media (Twitter and Facebook, for example) or commercial partners (Apple, for example) is prohibited by the dated terms of use in the 2009 Agreement – as is re-using providers' metadata in resources such as Wikipedia that can help aid public discovery but rigidly require liberal intellectual property licences.¹⁷ Key terms of the 2009 Data Exchange Agreement, in short, serve to wall off Europeana from the key parts of the web.¹⁸

These imperatives encourage Europeana and participating institutions to review the 2009 Data Exchange Agreement. One of the most important changes in the new Agreement is that it calls for a more open licence to govern the metadata held in Europeana's repository. The preferred licensing solution for the new Agreement is Creative Commons Zero (CC0), a universal public domain dedication.¹⁹ CC0 is quickly becoming the number one international standard for dedicating data and creative works to the public domain. As Creative Commons defines the licence:

¹⁵ http://ec.europa.eu/information_society/index_en.htm;
http://ec.europa.eu/information_society/policy/psi/index_en.htm;
http://ec.europa.eu/information_society/policy/psi/rules/eu/index_en.htm

¹⁶ http://ec.europa.eu/information_society/activities/digital_libraries/doc/letter_1/index_en.htm;
<http://en.wikipedia.org/wiki/Europeana>

¹⁷ "Your Metadata and Europeana";
http://en.wikipedia.org/wiki/Wikipedia:Copyrights#Reusers.27_rights_and_obligations and
<http://creativecommons.org/licenses/by-sa/3.0/legalcode>

¹⁸ Europeana is one of many cultural institutions coming to terms with this truth. Initiatives are under way
<http://wiki.okfn.org/OpenDataLicensing>;
http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf;
<http://bit.ly/qBEllE>.

¹⁹ CC0 is the most open tool Creative Commons offers. By applying CC0 the rights holder permanently waives copyright and (if applicable) database rights to descriptive metadata, granting the data public domain status. For more information on CC0 see <http://wiki.creativecommons.org/CC0>



CC0 enables scientists, educators, artists and other creators and owners of copyright or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

In contrast to CC's licenses that allow copyright holders to choose from a range of permissions while retaining their copyright, CC0 empowers yet another choice altogether – the choice to opt out of copyright and database protection, and the exclusive rights automatically granted to creators – the “no rights reserved” alternative to our licenses.

5. Business Model Perspective

During the workshop in July 2011 Europeana focused on the consequences of releasing metadata under CC0 for the business model of cultural institutions.

“If we do nothing and stick to traditional activities we will become invisible”

The main questions we sought to answer were formulated as following:

- What is the potential impact to your business model if, as a metadata provider, you start to release your metadata under CC0?
- What are the main potential benefits and risks of releasing metadata under CC0?
- What can we do to overcome the risks and start reaping the benefits?

5.1. The role of metadata in current business models

When discussing the impact on business models of making metadata available openly, the first thing one ought to know is what role metadata plays in current business models. Although the actual (strategic) role of metadata may differ from one provider to another, we suggest that three basic types can be distinguished.

1. Metadata as a Key Activity – Public Mission

One of the core activities for museums, libraries and archives is usually to create and maintain descriptive metadata. This can therefore be seen as an integral part of the activities that are needed to operate a heritage institution. As such it does not need to be related either directly or indirectly to the revenues that the institution generates. It is an integral part of the institution's public mission.

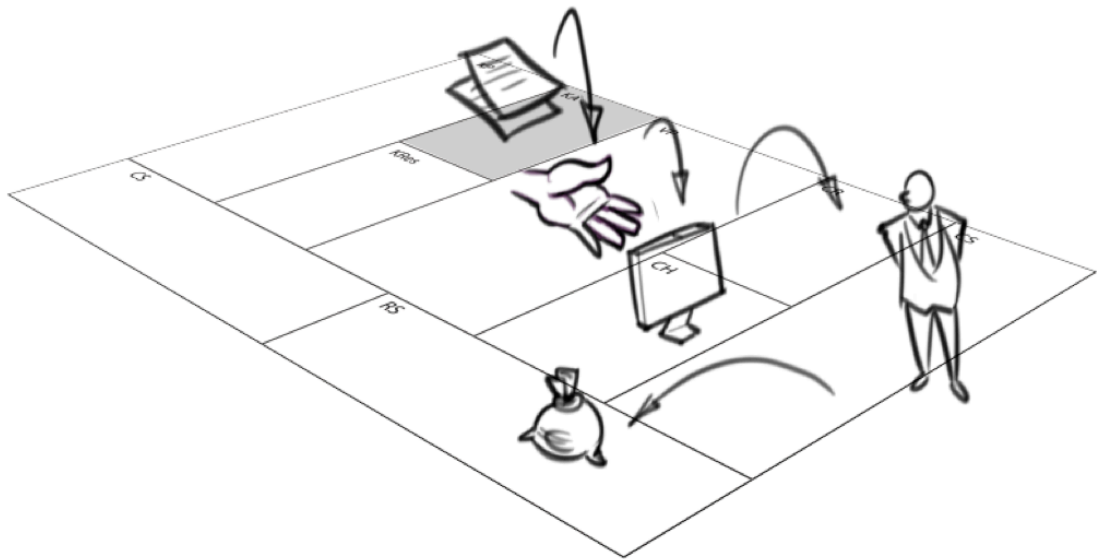


Figure 4: Metadata as a Key Activity

On the Canvas this can be depicted as a 'back-end' activity: metadata is created by the organisation and made available to the public. The client in this case is the government who in turn funds the organisation.

2. Metadata as a Key Resource – Indirect Revenue Stream

Metadata can also be of indirect importance to the income generated by a cultural heritage institution. In this case metadata should be seen as a strategic resource that is important (if not vital) to realise or maximise revenues from other value propositions. For instance, metadata can be used as a promotional tool for the actual content (books, magazines). It is thus used for marketing and branding purposes, both of which are important to realise sales of high resolution images or to attract more people to the institution or website. If metadata is used indirectly to generate revenue, it is no longer just an activity in business model terms. Instead it has become a key resource.

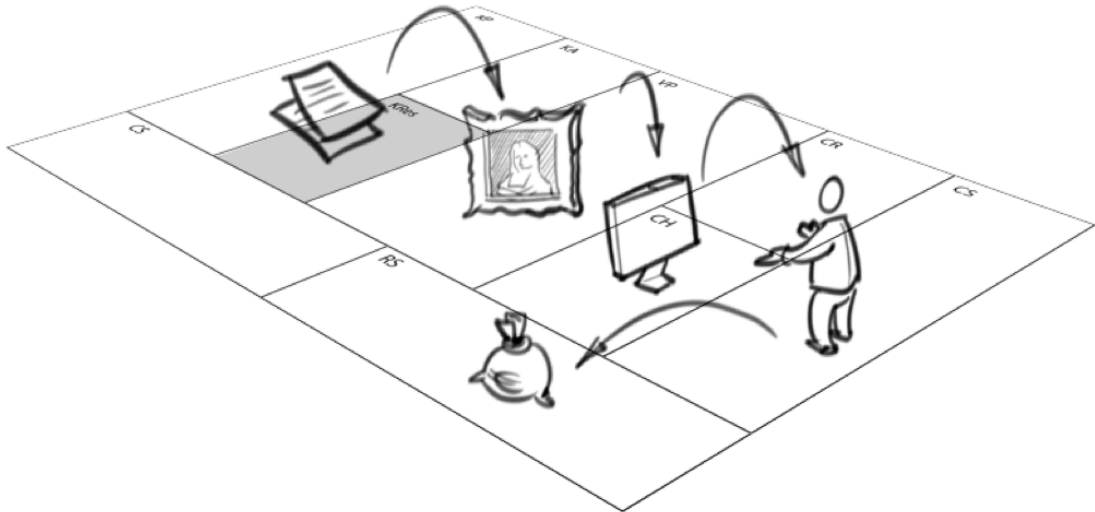


Figure 5: Metadata as a Key Resource

Metadata is used as a support mechanism for the value proposition, namely reaching a wider audience for the object by advertising, and if made available through the appropriate channels this will lead to income from users (professionals who need high resolution images for example, or individual end users who are prompted to visit the library or museum to view the original object).

“Metadata should be seen as advertisement for content.”

3. Metadata as a Core Value Proposition – Direct Revenue Stream

To some cultural heritage institutions (primarily national libraries) metadata is an important commodity they can use to generate direct income. If metadata is sold or licensed to other heritage institutions – or perhaps even to (professional) users – it is part of the core value proposition of the institution. For example, the German National Library (DNB) sells (tax-exempted) metadata to the library networks and other customers in Germany and elsewhere. In 2010, DNB began work on changing this business model, gradually moving towards providing its data free of charge for download, use, and processing. This is a first step in the process of taking a new perspective on the value of the services offered. Likewise the British Library recently released millions of records as Linked Open Data, under a CC0 licence²⁰. The British Library also sells metadata directly, which as they say is ‘worth millions’. They were able to take this step by creating a product/service differentiation, which allowed for

²⁰ <http://www.bl.uk/bibliographic/datafree.html>



the open publication of their data in RDF format, while commercially exploiting their full MARC 21 records.

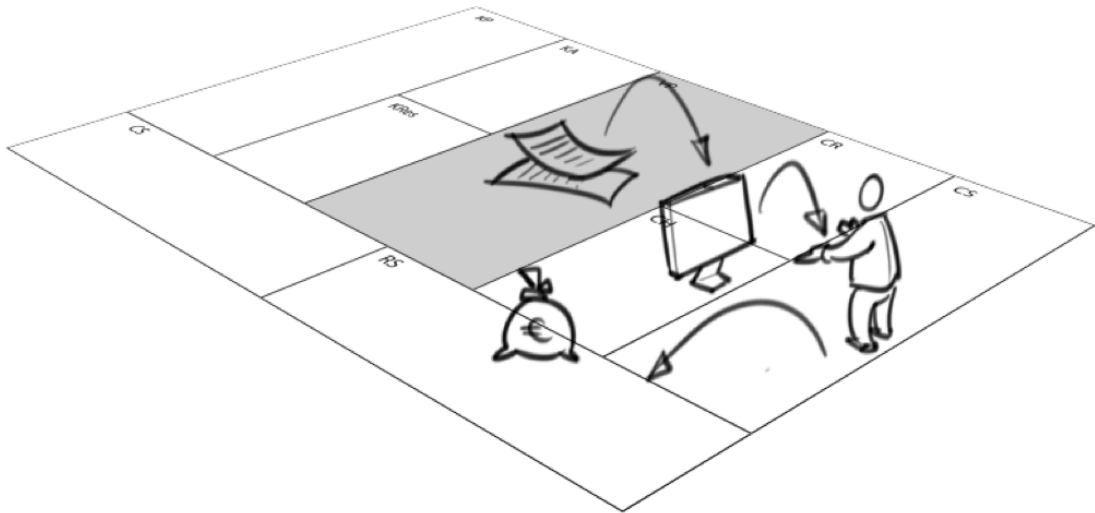


Figure 6: Metadata as a Core Value Proposition

Metadata in this model is the value proposition to a distinct (professional) user group willing to pay for this service.

“Most potential income should be seen as phantom income. But the fear of loss of this potential is very real.”

Only a limited number of institutions generate significant revenue by selling metadata directly to customers. But in an age where many cultural institutions are under severe pressure to generate income directly from the market this has become a potential revenue model many cling to.

These three types of value of metadata provide different starting-points for metadata providers seeking to change their business model through the release of open metadata.

5.2. Risks and benefits of innovating the business model

The potential benefits of open metadata



We distinguish ten major potential benefits:²¹

1. Increasing relevance: open metadata can be used in places where online users congregate (including social networks), helping providers to maintain their relevance in today's digital society.

2. Increasing channels to end users: providers releasing data as open metadata increase the opportunities that users have to see their data and their content.

3. Data enrichment: open metadata can be enriched by Europeana and other parties and can then be returned to the data provider. Opening the metadata will increase the possibility of linking that data and the heritage content it represents with other related sources/collections.

4. Brand value (prestige, authenticity, innovation): releasing data openly demonstrates that the provider is working in the innovation vanguard and is actively stimulating digital research.

5. Specific funding opportunities: releasing metadata openly will potentially grant providers access to national and/or European funding (European and most national governments are actively promoting open metadata).

6. Discoverability: increased use and visibility of data drives traffic to the provider's website.

7. New customers: releasing data openly offers new ways to interact with and relate to customers.

8. Public mission: releasing metadata openly aligns the provider with the strategic public mission of allowing the widest possible access to cultural heritage.

9. Building expertise: releasing metadata openly will strengthen the institution's expertise in this area, which will become a marketable commodity such as consulting services.

10. Desired spill-over effects: institutions and creative industries will be able to create new businesses, which in turn will strengthen the knowledge economy.

The potential risks of open metadata

1. Loss of quality: the high-quality metadata provided will be divorced from the original trusted source and corrupted by third parties.

2. Loss of control: institutions will no longer be able to control the metadata if anyone can re-use or distribute it.

3. Loss of unity: metadata will get scattered across the digital universe while it

²¹ The lists of benefits and risks were drafted from different consultation sessions organised by Europeana. The lists were then matched with insights from existing literature on the subject. Finally the lists were tested in the July workshop.



should be (contextually) kept together.

4. Loss of brand value: by releasing data openly the institution risks being associated with re-users that they do not want to be associated with.

5. Loss of attribution: by releasing data under an open licence institutions will not be credited as the source/owner of the metadata.

6. Loss of income: institutions are afraid that they cannot replace current revenues from metadata with other sources of income.

7. Loss of potential income: in the future institutions may think of a way to make money from metadata, but if they release it openly now someone else may do this.

8. Unwanted spill-over effects: institutions find it unfair that others make money with the metadata that they provide.

9. Losing customers: if data is openly available customers will go elsewhere to get the information they are looking for.

10. Privacy: there are privacy restrictions on the use of certain data.

‘We don’t necessarily want to make money ourselves, but why should others be allowed to do so based on our metadata?’

Weighing perceived benefits and risks

While a number of studies are available which describe the potential macro-economic effects of data, relatively little is known on the effects at the institution level. This paper therefore focuses mainly on the perception of heritage professionals of the benefits and risks associated with open metadata.

We asked the participants to rank the three most promising benefits as well as the three most threatening risks.

Cultural heritage professionals agreed that opening up metadata is vital in the long run to the **relevance of the institution** in modern, digital society. It was widely felt that this closely aligned with what they see as their **public mission to open up access** to our collective heritage. On a more technical level all could see the important role that open metadata can play in promoting access to the object to **new customers** by functioning as an advertisement for the object.

“A pilot carried out by the National Archive (UK) established that users trusted National Archive data over similar anonymous data 10 to 1.”

On the risk side there was more debate. Privacy rights of individuals whose names and materials are found in documents was mentioned often as a show-stopper, although this seems to be of particular importance to the archives and less so for museums and libraries. Opening up metadata could potentially be harmful to



individuals who will become more exposed to criticism. It should also be mentioned that in most jurisdictions privacy issues are dealt with in privacy laws (which may forbid disclosure of data under open conditions). As such the privacy issue is the only risk that cannot be tackled within the framework of a business model. Bearing this in mind, there was agreement that **loss of attribution, loss of potential income and unwanted spill-over effects** were seen as the most important threats of opening up metadata to the business model of the institutions.

From the debate that accompanied the selection exercise two principal conclusions can be drawn:

1. Time gap

First, there is a general feeling that in the longer run the perceived benefits will outweigh the risks. The benefit of 'increased relevance', for example, will grow over time (for instance, as new users start to use metadata in different contexts like social networks) and new customers are acquired. Since business models for open metadata are still in their embryonic phase, it is plausible to assume that potential benefits will be realised only in the long run.

The negative effects of releasing open metadata might be felt directly in the short run. CC0 and other open licences are irrevocable. The risk of loss of potential income is therefore very real to heritage institutions. For those few institutions that monetise metadata directly, opening up their metadata could immediately adversely affect their cash flow.

In short, heritage professionals believe there is a time gap between the negative effects of open metadata and the full realisation of its benefits.

2. The current role of metadata matters

The second conclusion is that the current role of metadata in the business model of a heritage institution matters when that institution is considering opening up its metadata.

Recall that we identified three basic types: one in which metadata is an activity that is not directly or indirectly related to the generation of revenues; one in which metadata is a key resource that is of indirect importance to revenue generation; and one in which metadata is the value proposition and therefore a direct income source. For each of these models we have tried to investigate the balance between benefits and risks that can be realised in the **short to medium term**.

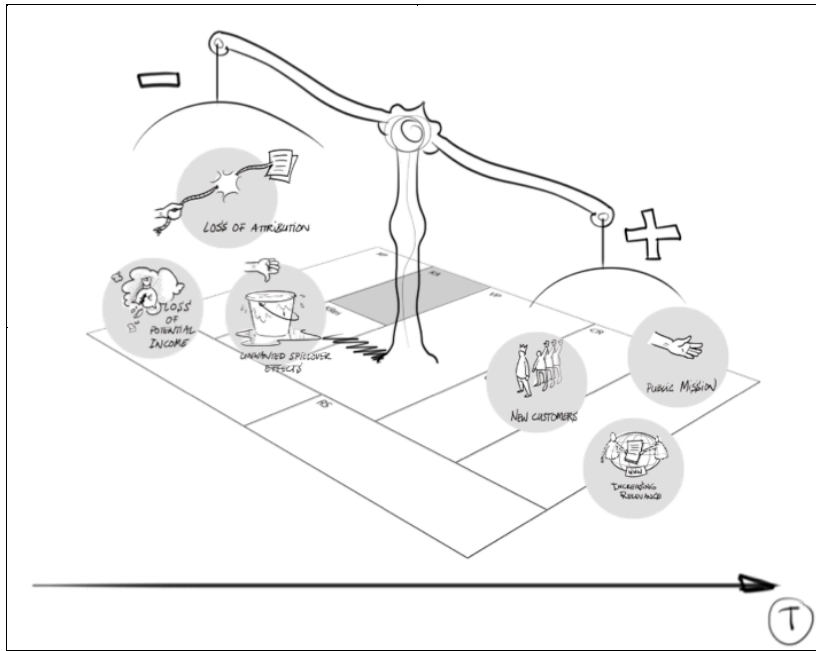


Figure 7: Medium-term balance for metadata as a key activity

When metadata is a key activity, which is part of the institution's public mission, the balance will quite certainly be positive after releasing metadata under CC0. The loss of potential income and unwanted spill-over effects are unlikely to be seen as important risks, as in this business model metadata does not play a role in revenue generation and spill-over effects are seen as a desired outcome. Loss of attribution may be an issue. On the side of the benefits new customers may not be so important. However, increasing relevance and public mission clearly are. From a business model perspective, there appears to be little to deter heritage institutions of this type from releasing metadata as open metadata.

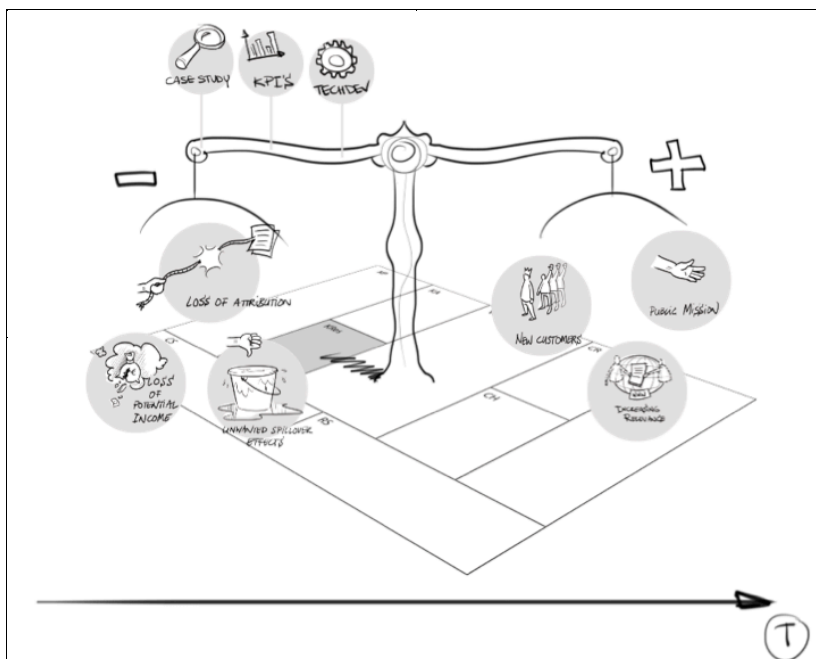


Figure 7: Medium-term balance for metadata as a Key Resource



The balance may pan out slightly differently for heritage institutions that use metadata as a key resource. Cultural heritage professionals seem positively inclined that in the long term the balance will shift in favour of the benefits. Opening up metadata should naturally result in increased visibility of their assets that can be reached through many new channels and result in new customers. However, it is felt that this will only work when material is properly attributed. Otherwise the fear is that this shift may result in negative effects, such as of loss of income and unwanted spill-over effects ('others could make money on 'my' metadata').

“Every institution should be curating its assets on Wikipedia. It should be part of every business plan.”

Some solutions to this problem have been mentioned. Three are especially relevant to mention here. First, there is an urgent need to bring up success stories – **case studies**, for example, illustrating how things work out in practice (see the Appendix). Second, the perception of risks (and of benefits) is closely related to how we measure and validate success. Re-examining **key performance indicators (KPIs)** and underlying funding mechanisms will be helpful. For instance, metadata is still used on a large scale to attract people to an institution's website. From that perspective releasing metadata openly can easily be seen as a threat as it may pull visitors away to other online places. Yet it might make more sense if KPIs were to measure the effect of distribution of institutional information to places online 'where users want to be'. In other words, instead of stimulating a gravitational approach whereby users are attracted to one particular website, the KPI should encourage access generation at sites already frequented by users (for instance social networks). Third, **technological solutions** can be thought of (and are in fact already being implemented) to keep track of metadata in the network so as to prevent a loss of attribution and to signal new business opportunities.

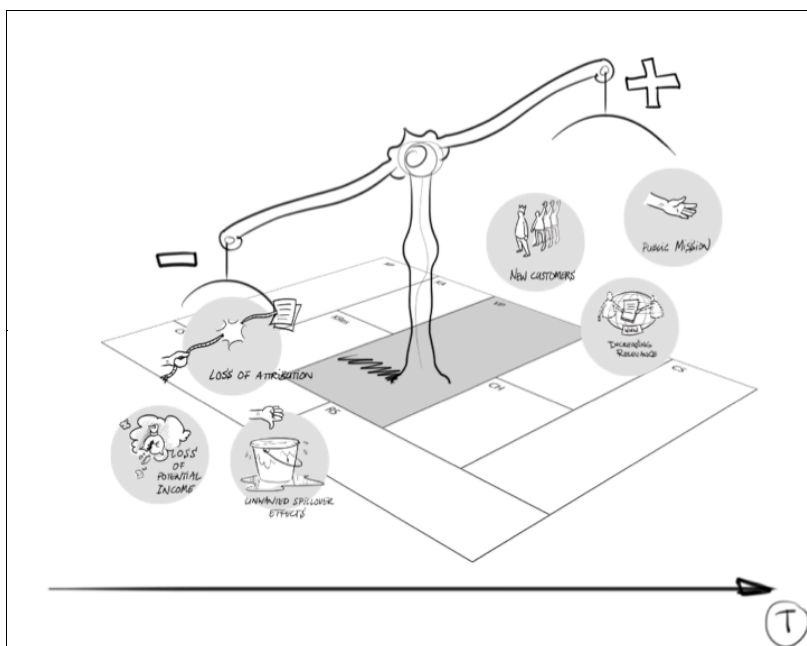


Figure 8: Medium-term balance for metadata as a Key Value Proposition



In the short and medium term there is a perceived high risk for those cultural heritage institutions that make direct income selling metadata. Releasing metadata under CC0 immediately challenges their current business model with the loss of potential income for the institution and unwanted spill-over effects on the management's horizon. It has been argued that the loss of this income can be averted by product differentiation: data can be made available openly in RDF format so that it becomes suitable Linked Open Data, while full MARC 21 records can still be marketed under commercial terms. A larger issue is the fear of losing the opportunity to sell data in the future when data is openly available for everyone to use. This requires a change of mindset and the acknowledgement that the reality of the web in the 21st century is that we are all invited to create new, commercial services based on open data.

An important side note is that this model applies to a very small minority of heritage institutions (mostly national libraries) and appears to be of decreasing importance, as the British Library is showing. Furthermore, one essential thing to keep in mind is that in the context of Europeana, release under CC0 is required only for descriptive metadata. Heritage institutions may very well **split metadata** between basic descriptive metadata and research-rich metadata that is used as a value proposition.

6. Conclusion

“The risk is not to be there when the public needs us most. So we should play a prominent role in this space, not lose it to anyone else.”

As one of the workshop professionals observed, the single most important risk that cultural heritage institutions run is to miss out on the digital transition that is reshaping society.

We recommend that three specific issues need to be addressed:

1: Spill-over effects: Opening up data should be seen (again) as an important part of the *raison d'être* of our public cultural sector. Instead of measuring success by the amount of commercial revenue that institutions are able to secure from the market, new metrics should be developed that measure the amount of business developed (spill-over) based on data made openly available to the creative industries. This requires a change on a policy level.

2: Loss of Attribution: Heritage institutions are the gatekeepers of the quality of our collective memory, therefore a strong connection between the object and its source is felt to be desirable. There is a fear that opening up metadata will result in a loss of attribution to the memory institution, which in turn will dilute the value of the object. Investigations need to be made on the technical, legal and user levels to safeguard the level of integrity of the data.

3. Loss of potential income: It has been established that a very limited amount of Institutions currently make significant money selling metadata. It must be argued that the loss of this income can be averted by product differentiation. A larger issue is the fear of losing the **opportunity** to sell data in the future when data is openly available for everyone to use. This requires a change of mindset and the acknowledgement that



the reality of the web in the 21st century is that we are all invited to create new, commercial services based on open data.

Overall we can conclude that there is a strong conviction among cultural heritage professionals that the benefits of open sharing and open distribution will outweigh the risks. In most cases the advantages of increased visibility and relevance will be reaped in the short term. In other cases, for example where there is a risk of loss of income, the advantages will come in the longer run and short-term fixes will have to be found. All of this requires a collective change of mindset, courage to take some necessary risks and a strong willingness to invest in the future of the society we serve and participate in.

7. Literature

- Creative Commons (2011), *The Power of Open*, available at <http://thepowerofopen.org/>
- Deckers, Polman, te Velde, de Vries, *Measuring European Public Sector Information Resources*, available at <http://bit.ly/qSryyi>
- European Commission (2010), *Unlocking the Potential of Cultural and Creative Industries*, available at <http://bit.ly/bgBcoG>
- European Commission (2011), *Open Data and the re-use of Public Sector Information, report from the workshop Digital Agenda Assembly 17-17 June*, available at <http://bit.ly/gLslcA>
- Hargreaves (2011), *Digital Opportunity: A review of Intellectual Property and Growth*, available at www.ipo.gov.uk/ipreview-finalreport.pdf
- IRN research (2011), *Europeana Online Visitor Survey*, available at <http://bit.ly/nTgxf6>
- Kroes, Neelie (2011), *From Closed Doors to Open Gates*, Uncommon Culture, Vol 2, no. 1(3) available at <http://bit.ly/sgW4Mw>
- Niggemann, De Decker, Levy (2011), *The New Renaissance, report by the Committee des Sages*, available at <http://bit.ly/g5RjCn>
- Osterwalder, Pigneur (2009), *Business Model Generation*, available at <http://www.businessmodelgeneration.com/>
- Weber, Steven (2004), *The Success of Open Source*. Cambridge, Harvard University Press
- Von Hippel, Eric (2005), *Democratizing Innovation*. Cambridge, MIT Press. <http://web.mit.edu/evhippel/www/books.htm>



8. Appendix: Case studies

In order to build the evidence base it is helpful to collect examples from cultural and educational institutions that are opening their digital doors to this kind of innovation for (broadly defined) business –reasons as well as for mission-oriented reasons regarding the diffusion of knowledge. With this objective, we have initiated a set of ongoing case studies of institutions that are engaged in opening their data and metadata, even in the strict sense of the word, for business reasons. This includes major institutions – the British Library, the British Museum, Cambridge University, MIT, Sound and Vision, Yale – plus commercial publishing houses and broadcasters in Europe and abroad that are funding or participating in the linked open data movement and expecting substantial returns on investment for doing so. The first results of this information gathering are published with this White Paper. These case studies have been collected through interviews with key decision makers over the summer of 2011.

Cambridge University

Among the cultural and educational institutions leading the charge into the open data movement is Cambridge University Library, which initiated the Cambridge Open Metadata (COMET) project with funding from the UK government agency JISC. Deliverables from this initiative (to be found at: <http://data.lib.cam.ac.uk/>) include approximately 1.5 million bibliographic records that Cambridge University itself compiled and created and 1.7 million bibliographic records (in process) from vendors who initially provided those records to the British Library and Research Libraries UK, the open licence to which (<http://opendatacommons.org/licenses/pddl/1.0/>) COMET secured from those same vendors. In addition, COMET compiled a guide to who owns metadata – bibliographic metadata in particular, and MARC 21 records to be precise. This guide, online at: <http://cul-comet.blogspot.com/p/ownership-of-marc-21-records.html> is remarkably useful for those who seeking to appreciate the complexity of metadata rights and licensing in 2011.

In an August 2011 interview, **Ed Chamberlain** at Cambridge University Library described two sets of pressures coming to bear on the library that helped to give rise to the COMET project. The first set involved academics – including those close to the Open Knowledge Foundation (<http://okfn.org/>), based at Cambridge – who asked why taxpayer-funded and thus publicly-supported data was still being siloed within the university and effectively kept from public access. The second set involved technologists and librarians who wanted to see records becoming part of the linked open data movement, where data can be made available for general discovery and where such an approach, in Chamberlain’s words, “heightens the chance that someone can do something interesting”. For Cambridge, which will continue to make more such data available under liberal licences, the eye is very much on web trends that take advantage of exposed data, even though cases of use from the wider community are, in Chamberlain’s words, still “perhaps 18 months away” from being able to be documented.

British Library



In quantitative terms the British Library (BL) currently leads all libraries, having provided approximately 2.6 million of 14 million catalogue records in the BL's National Bibliography (<http://www.bl.uk/bibliographic/natbib.html>) – with the rest soon to follow (<http://www.bl.uk/bibliographic/datafree.html>). Interestingly the BL decided to release this data subset, covering books published or distributed in the UK since 1950, even as it sells commercial versions of the same datasets to customers worldwide; it “remodelled” its MARC bibliographic data for XML-based RDF delivery, which it has provided to the public for free under a most liberal Creative Commons Zero licence (<https://creativecommons.org/about/cc0>).

For the BL's **Neil Wilson**, several factors were converging to instigate this new policy – from existential concerns about the relevance of libraries to concerns about their siloed and dated approaches to data. There was interest in what the scientific community was doing with linked open data and a sense that the Library could participate in that kind of excitement online. “Imagine”, Wilson said in an August interview, “relating resources in different domains – linguistics, geography, political science, for example – to the field of book data, and learning more as a result about who wrote a particular book, where, and why”. The Library had already been engaged in controlled experiments with datasets - one with philosophy books mapped out against time and place of publication, and another with the UK's Intellectual Property Office, looking through 5 or 6 million records and 400 years of copyright legislation to help see how copyright legislation affects artistic creation not only with books but music and moving images as well. Also driving this movement was the UK government's evolving policy – across political parties – requiring the majority of government-published information to be re-usable linked data in 2011.

The BL was able to affect this shift in policy because of economic trends in the library world – especially libraries being asked to move to new levels of efficiency by outsourcing the creation of more data to speciality vendors that sell that data to institutions. The BL had a whole unit processing and reselling that data; as Wilson says, “because we licensed it to sell, we were able to give it away” as well. The BL was also expecting a decline in the sales of its MARC bibliographic data in a world where more and more books are being distributed in digital form. Although the pilot is only weeks old as we go to press, Wilson believes that an “ecosystem of sorts” is evolving, one where linked data will enable all sorts of discovery and applications, commercial and non-commercial, in the months and years ahead.

JISC

While Cambridge might be the most aggressive institution in sorting through the business and legal complexities of metadata rights and ownership, and the BL the most aggressive in pushing its data toward open, the UK government higher education agency JISC has been building platforms for conversation - at least in the UK - for some years. These include manifestos regarding the benefits of open content and open linked data, lists of signatories supporting the same (<http://discovery.ac.uk/businesscase/signatories/>), and a detailed index of progress on this front among major cultural and educational institutions (<http://obd.jisc.ac.uk/examples>).²²

²² <http://ckan.net/tag/library>



For **Andy McGregor** from Discovery, the JISC-funded programme that aims to create “a thriving metadata ecosystem”, speaking in an August interview, the benefits of linked open data will become much clearer in the next year or two. Later in 2011, JISC will sponsor concentrated public-private initiatives around linking open data regarding the works of William Shakespeare and the centenary of World War I, among other subjects. Meanwhile, JISC and others plan to 1) collect metrics around the developments the commercial sector is starting to make with open data; 2) list more “shining examples” of the best-use cases, such as the BBC’s Wildlife Finder (<http://www.bbc.co.uk/nature/wildlife>; <http://ckan.net/package/bbc-wildlife-finder>) and HistoryPin (<http://www.historypin.com/>); and 3) locate the best spokespeople to get behind open data and say publicly how useful it is. Competitions to work with open data published with the support of JISC Discovery and new hackathons are part of JISC’s 2011 and 2012 plans.

Yale University

In May 2011, Yale University announced its new open access policy governing the digitisation of the millions of objects held in the university’s own museums, archives, and libraries. Specifically, Yale declared that no licence would be required for the use of the high resolution images of those works that are in the public domain, that no legal or business restriction would govern the transmission of these images, and that no limitations would or should be imposed on their use. The intended result: “that scholars, artists, students and citizens the world over will be able to use these collections for study, publication, teaching and inspiration.”²³

According to Yale’s **Meg Bellinger** in an August 2011 interview, Yale’s new policy required two years to develop, debate, and finally articulate. Among the accelerants: museum directors keen on exposing their work to more people; technologists enthralled to see what the public might do with this information; recent (U.S.) legal decisions finding that digitisation is not infringement; and funders (including private foundations) increasingly concerned with the rising costs of art publications and the costs of licences to certain images in particular. In the three months since the new policy was declared, Bellinger says, Yale has heard from “a lot of happy people” – instructional technologists, average citizens (some of whom have corrected the university’s metadata), and publishers who find it easier now to discover and use artwork images in particular. While the public has been enthusiastic about using “cross collection” discovery portals that Yale has built – <http://discover.odai.yale.edu/ydc/> – the institutions at Yale are also finding new methods of cooperation between themselves. The Yale Center for British Art, for example, is harvesting metadata related to its collections in the university library’s book catalogue.

Yale established this new policy in 2011 notwithstanding a vigorous licensing and publishing programme associated with its image collections. Arguments concerning the opportunity cost of open access (giving away potential revenues, for example) are based less on specific examples than on hypothetical opportunities – “the magic app” – that frankly never materialise. The university has found that those publishing partners and licensors who want to resell Yale cultural heritage content generally are interested in reselling Yale’s brand – the university’s name, logo, and other trademarks

²³ <http://news.yale.edu/2011/05/10/digital-images-yale-s-vast-cultural-collections-now-available-free>



- and that brand remains under strict licensing provisions.²⁴ Open access policies and licensing programmes can coexist well especially when the brand is key to a licensing programme. Indeed, Bellinger reports numerous inquiries into commercial licences that have appeared in the wake of the publicity generated by Yale's open access declaration.

British Museum

The British Museum has also taken a major policy decision in September 2011 by releasing sets of its data in the W3C open data standard RDF in order to, as the policy states, allow the Museum's collection data to "join and relate to a growing body of linked data published by other organisations around the world interested in promoting accessibility and collaboration" (<http://collection.britishmuseum.org/Licensing>). Here, a lot of the momentum is being carried by the UK government, which is establishing new policies for its own data – see: <http://data.gov.uk/> – with the aim of "opening up" government. But in recent years the Museum has found collaborations with institutions that use and indeed regularly harvest the Museum's metadata to be helpful and informative for the Museum's own mission. UCLA's Cuneiform Library (<http://cdli.ucla.edu/>) is one case in point, and the Connected Histories project (<http://www.connectedhistories.org/>) is another.

The Museum's **Dominic Oldman**, in an August interview, stressed that when the Museum debates open access policies, any downside of the inappropriate use of data is completely outweighed by the benefits of proper re-use that facilitate scholarly research and public discovery. In addition, the Museum is continually interested in exploring how its data is being used with a view toward updating its own data policies and services. The non-commercial clause that has governed use and re-use of the Museum's metadata is rooted in the belief that non-profit academic charities should enable free use only for non-profit purposes. But in the digital age, with evidence that use and re-use can increase knowledge when it is openly linked across the entire web, the new view is that data funded by the taxpayer should have the broadest possible distribution. In addition, executives at the Museum reportedly believe that many of the significant commercial activities of the institution through its British Museum Company division could be enhanced through greater exposure online.

German National Library (DNB)

In 2010 the German National Library (DNB) began work on, as its website puts it, "completely changing its business model". The library's ultimate objective at the time was, and remains, "providing its data" – data it was generating, data it was acquiring from third parties – "free of charge for downloading, using and processing." The DNB previously had provided bibliographic data freely for research purposes, but it was used to charging customers and clients for that data if that data was being in any way re-used.

The DNB is describing its new approach as in accord with the European Commission policies and proposals concerning publicly generated information. But, interestingly, the DNB goes further. According to information provided by the DNB's **Jan Hannemann**, the DNB is also collecting information from new studies being

²⁴ <http://www.yale.edu/licensing/>; http://www.wipo.int/sme/en/documents/uni_trademark_licensing.html



conducted in the field: “Studies carried out on the changes which have taken place as a consequence of the EU regulations already enacted (PSI Directive - Re-use of Public Sector Information) have shown that the use and re-use of digital information have increased dramatically following the reduction or abolition of fees, that new users have been attracted and that innovative areas of re-use have been created, meaning that ‘the social and economic advantages far outweigh the short-term financial benefit of cost-effectiveness in cases where no or only low charges [...] are levied.’”²⁵

These institutions are not alone in embracing linked open data and open access practices. And indeed, many commercial companies that also work in the field of education and culture – publishing houses in particular – are finding that benefits from participating in linked open data outweigh the risks.²⁶ As Europeana builds its relationships with its data providers on the one hand, and the rest of the worldwide web on the other, its partners are likely to find that the broader the access they provide to their metadata, the better these benefits will be.

²⁵ <http://www.d-nb.de/eng/service/zd/geschaeftsmodell.htm>;

http://www.d-nb.de/eng/hilfe/service/linked_data_service.htm

²⁶ http://www.cni.org/topics/identity-management/yaleimages-id-data_reuse/