DELIVERABLE

Project Acronym	**Linked Heritage**
Grant Agreement number	**270905**
Project Title	**Coordination of standard and technologies for the enrichment of Europeana**

Addendum to D4.1 Best practice report – Public–Private Partnership

Revision: 1.0

Authors:

Michael Hopwood (EDItEUR)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	
C	Confidential, only for members of the consortium and the Commission Services	X

Revision History

Rev	Date	Author	Organisation	Description
0.1	Feb 2013	Michael Hopwood	EDItEUR	First draft
0.2	May 2013	Michael Hopwood	EDItEUR	Second draft after internal (EDItEUR) review
0.3	June 2013	Michael Hopwood	EDItEUR	Final draft for internal review
0.4	July 2013	Graham Bell	EDItEUR	Draft for work group review
1.0	Aug 2013	Graham Bell	EDItEUR	For submission
1.1	01 Sep 2013	Claudio Prandoni	Promoter	Formal check

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY 4

2. INTRODUCTION 5

1.1 Background 6
1.2 Role of this Deliverable in the Project 6
1.3 Approach 6
1.4 Structure of the document 6
1.5 Europeana and Linked Heritage Ecosystem 6

2. PROBLEM OUTLINE OF PPP BUSINESS MODELS 8

3. PUBLIC–PRIVATE PARTNERSHIPS IDENTIFIED IN D4.1 9

3.1 Europeana materials for purchase 9
3.2 Classic digital library interfaces 11
3.2.1 Business models will be limited in the current framework 11
3.2.2 There is still scope to expand digital library content 11
3.2.3 A new framework could enable new models 12
3.2.4 The Centrality of Identification 12
3.3 Linked data business models? 15
3.3.1 Commercial product data is already "linked" data 15
3.3.2 Linked open data could work, but only with a licence 16
3.3.3 Proposed linked data business models 18
3.4 Commercial Data Aggregations 21
3.4.1 Commercial sector partnerships will be culturally valuable 21
3.4.2 Metadata's value derives from content's value 21
3.4.3 Existing cultural content licensing for commerce 22

4. ALTERNATIVE TYPES OF PARTNERSHIPS 23

4.1 Reusing cultural heritage sector data 23
4.1.1 Survey of existing re-uses 24
4.1.2 Survey sources and method 24
4.1.3 Survey results 25
4.1.4 Types of application and API implementation 25
4.1.5 Types of development organisation 25
4.1.6 Patterns of product development 26
4.1.7 Patterns of reuse and recombination 26
4.1.8 Retail availability of apps 27
4.2 Collaborative data publishing 27

5. CONCLUSIONS 30

5.1 Results From survey of apps reusing Cultural heritage data 30
5.2 Likely sources of bulk metadata from the commercial sector 30

6. APPENDIX – SURVEY SAMPLE 31

1. EXECUTIVE SUMMARY

This addendum to the main D4.1 *Public–Private Partnership: Best Practice Report*[1] expands upon the information therein. It restates and emphasizes that for true partnership with the commercial sector, the costs of participation are not trivial, and Europeana must deliver benefits to that sector that outweigh the cost of participation.

In particular, the release of commercially-relevant metadata into a CC0 rights-waived data environment driven by Europeana’s Data Exchange Agreement poses challenges to a commercial sector where there is already a lively commercial market for rights-controlled metadata. The removal from Europeana of formerly-available metadata describing commercially-available products following imposition of the CC0 provides some evidence of this. Anecdotally, the loss of rights and control over the contexts in which the metadata may be reused may weigh more heavily than the loss of direct commercial value of the data.

The report describes an informal survey of past efforts to generate value from reuse of the existing Europeana metadata, which mostly derive from ‘hackathon’-style events and project-supported development rather than commercial investment, and concludes that an insignificant proportion of these developments result in commercially-relevant products. The relatively low ‘viewership’ of Europeana (each object page within Europeana is viewed much less than once per year, on average) limits the opportunities for value generation, and inconsistent identifier policies in the cultural sector make it difficult for commercial organisations to reuse the Europeana metadata for enrichment or enhancement of existing commercial metadata.

Two potential developments may provide solutions: first, the adoption of rights frameworks other than CC0 might provide the reassurance over context that commercial organizations desire, and encourage the delivery of more metadata records to Europeana; second, providing increased value to the commercial sector through facilitating licensing and reuse of the *digital cultural objects themselves* (rather than simply facilitating their discovery) through a simple licensing service. Metadata’s value derives from the value of the content it describes, and easing the commercial licensing for reuse of that content bolsters the value of the service that Europeana provides.

The report also considers which commercial organisations might be most receptive to future metadata collaborations, either for provision of data to Europeana or for reuse of Europeana metadata to enrich existing commercial data. It concludes that the most likely sources of accurate, bulk metadata are the various Books in Print and commercial data aggregators, rather than publishers or retailers. However, these aggregation services have no direct interest in either greater reuse of the data or in greater sales or licensing of the underlying cultural object (where the benefits accrue to the publisher and retailer). Thus partnerships, between the data aggregators and the publishers that provide original metadata to them, are likely to be required. The direct engagement of publisher-owned data aggregator, Libreka, with Europeana lends credence to this.

[1] Available from http://www.linkedheritage.eu/getFile.php?id=283

2. INTRODUCTION

Europeana has built a Europe-wide cultural heritage service comprising several key aspects:

- The Europeana.eu discovery **portal**;
- **Textual metadata** that supports the portal, available to the public through
  - APIs
  - Linked Open Data sets;
- **Preview images** hosted locally at Europeana's servers;
- The substantial **supporting network** of data contributors, project partners and development experts that surround all of these.

Until recently all of these aspects have focussed on the *public sector* (publically-funded) heritage activities of Europeana's network.

However, attention has turned to interaction with the *commercial sector*.

One year before the start of Linked Heritage, in June 2010, the Federation of European publishers (FEP) publically stated:

> *"Successful projects at national level such as **Gallica** in France, **Enclave** in Spain and **Libreka** in Germany could be furthered at EU level or inspire other projects.*
>
> ***Links through Europeana to sites of publishers or other rightsholders to enable users to buy in-copyright are the ideal way to achieve a private sector involvement**".*

This was the approach assumed by the EU publishing sector generally, and by EDItEUR on commencing our work with Linked Heritage.

The lifetime of the Linked Heritage project (2011 – 2013) saw more developments:

- The European Commission's *New Renaissance* report
  - recommended digitising in-copyright (but out of distribution) materials and making them available online;
  - specifically recommended *"to make in-copyright content offered by private providers against payment searchable through Europeana"* and said that *"Partnerships between Europeana and private companies active in online distribution (e.g. publishers) should be encouraged."*
- Linked Heritage Work Group 4 worked on the latter recommendation and produced three reports laying out the technical and legal-commercial specifications for including product metadata about commercial books, film and TV, recorded music and photography in Europeana;
- Work within Linked Heritage WG4 was compromised by the announcement of Europeana's DEA in September 2011[2], because of conflicts between the new CC0 terms and the perceived commercial value of rich metadata;
- Europeana has embarked this year (2013) on two projects that explicitly set out to engage the commercial sector in the Europeana network:
  - Europeana Creative (building pilot partnerships with commercial entities);
  - Europeana Cloud (providing a licensing model for Europeana Creative).

Within this context, and a the express wish of the EC project reviewers, this Addendum to Linked Heritage D4.1, attempts to address concerns and make clarifications about public–private partnerships with Europeana, writing from a commercial sector point of view informed by WG4's now completed exploratory work with international standards bodies and high-level commercial contacts in the media industries.

[2] For the full timeline, see http://pro.europeana.eu/consultation-process/

1.1 BACKGROUND

At the first project review of Linked Heritage in 2012, the EC's reviewers commented that

> *"The public–private partnerships identified* [in D4.1] *are all ways in which the cultural heritage material leads to a commercial front-end, either through a classic digital library interface or through linked data which can support this kind of business model.*
>
> - *In Europeana, there are already material*[s] *that can be purchased.*
> - *Alternative types of partnerships should be investigated."*

They requested that WP4 should

> *"investigate different types of partnerships that can be set up and present the potential of such partnerships to various industries, including models for*
>
> - *collaboratively publishing data*
> - *or reusing data from the cultural heritage sector."*

This Addendum addresses the reviewers' comments directly, aiming to give a more complete picture of the current situation and respond to these concerns from a commercial perspective.

1.2 ROLE OF THIS DELIVERABLE IN THE PROJECT

The main Tasks for WP4 given by the Linked Heritage Description of Work have been addressed in the three deliverables.

However, this Addendum provides extra context for Tasks:

- T4.1 Private Sector Business Exploration;
- T4.2 Contribution Specification.

Therefore it draws somewhat on the *business models and legal / licensing frameworks* described in D4.1 and D4.3. The *technical specifications* in D4.2 are relatively independent from these aspects, although not entirely separate.

This Addendum should not impact significantly on the final stages of Linked Heritage's other Work Packages; however, it is hoped that it will contribute to the work of

- Europeana Network's current PPP task force (and future work in this area);
- The two Europeana-partnered projects mentioned in 1 sopra, and future PPP projects.

1.3 APPROACH

As usual in EDItEUR's work on Linked Heritage, a combination of *documentary research* and *new data gathering* was used. Much of the background to this Addendum refers to the three complete Linked Heritage deliverables (D4.1–4.3). **A new survey of apps built on Europeana's and other heritage datasets was undertaken (see section 4.1.1 below).**

1.4 STRUCTURE OF THE DOCUMENT

The chapters which composing this document follow the EC reviewers' recommendations, answering each point in order. As background to our responses we preface an outline of the problem, based on that in D4.3.

1.5 EUROPEANA AND LINKED HERITAGE ECOSYSTEM

The current arrangements between data contributors, Linked Heritage and European are summarised in the diagram below, which also supplies the terms used in this report (CHO, DO, image preview *etc.*).

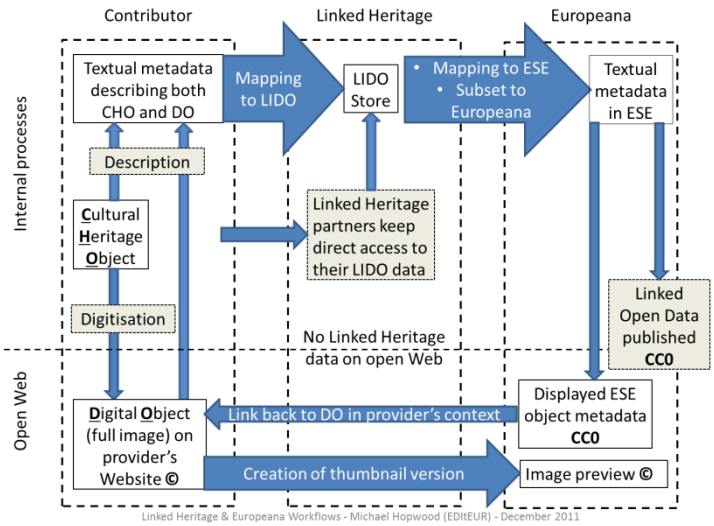


Figure 1 - workflows and key terms in Europeana / Linked heritage ecosystem

2. PROBLEM OUTLINE OF PPP BUSINESS MODELS

As outlined in D4.3, Europeana's engagement with the private sector can take one of two forms of fundamental business model:

1. The Linked Heritage specifications, based on the 2010 FEP statement and comprising commercial and technical opportunities:
   - A real commercial *quid pro quo*, for example through
     - Commercial links with retailers that enable click-through purchases, or
     - Provision of an identifier resolver that indirectly enables click-throughs;
   - Provision of data management tools that allow data suppliers to actively manage their dynamic data and be assured of the quality of the presentation of their data to end-users.

For more details of these opportunities, please refer to D4.3 and D4.2 respectively.

2. The second fundamental business model envisages that commercial entities might begin to make commercial use of Europeana's data (possibly in return for contributing some of their own).

This type of model goes somewhat beyond the remit of Linked Heritage, and rests on future developments in Europeana's core service and wider ecosystem:

- **Problem**:
  - Europeana simplifies *initial discovery* of digital cultural heritage objects and a *personal research approach to navigating* among them; it does this by aggregating the metadata in a format suitable for full-text search and subsequent display; it also makes this metadata available for reuse;
  - However, discovery, research and display of this sort is *not the primary problem* that commercial reusers face; commercial data services and metadata standards already enable this for some applications (for example, commercial picture research and licensing of fine art images), but they are not directly used by most cultural institutions (or Europeana itself);
  - Commercial reuse is difficult and expensive because of the disparate licensing frameworks and processes that the cultural institutions put in place for their original Cultural Heritage Objects and Digital Objects that underlie the metadata.
- **Possible solutions**:
  - Europeana would ideally provide tools required to enable *commercial reuse of the Cultural Heritage Objects and Digital Objects themselves*, for example:
    - a licensing framework for the DOs;
    - click-through licensing agreements;
    - centralised access to the DOs;
    - collective licensing schemes.
  - Pure metadata services based on Europeana's ecosystem of data providers could enable a *quid pro quo* entirely in terms of mutual data enrichment; this would rest on:
    - access to the richer metadata produced at the intermediate (aggregator) level of the ecosystem;
    - assurance of the consistency and standardisation of this data;
  - licensing frameworks *etc.* as for DOs and CHOs above, but in this case, for the licensing of large and rich datasets, and syndication of rich descriptive data covered by copyright.

We offer here some extra points to aid in future projects addressing this latter business model.

3. PUBLIC–PRIVATE PARTNERSHIPS IDENTIFIED IN D4.1

As the reviewers stated, the "*public–private partnerships identified* [in D4.1] *are all ways in which the cultural heritage material leads to a commercial front-end, either through a classic digital library interface or through linked data which can support this kind of business model.*"

This is identified as business model 1 sopra.

Linked Heritage took this approach following

- the project Description of Work itself;
- the lead of existing commercial data providers;
- and also the Federation of European Publishers, a Europeana Network member, who, before the start of Linked Heritage, had already stated:

  "Links through Europeana to sites of publishers and other rightsholders to enable users to buy in-copyright content are the ideal way to achieve a private sector involvement."[3]

A simple diagram (*Figure 2* below) shows this process from the customer's point of view.

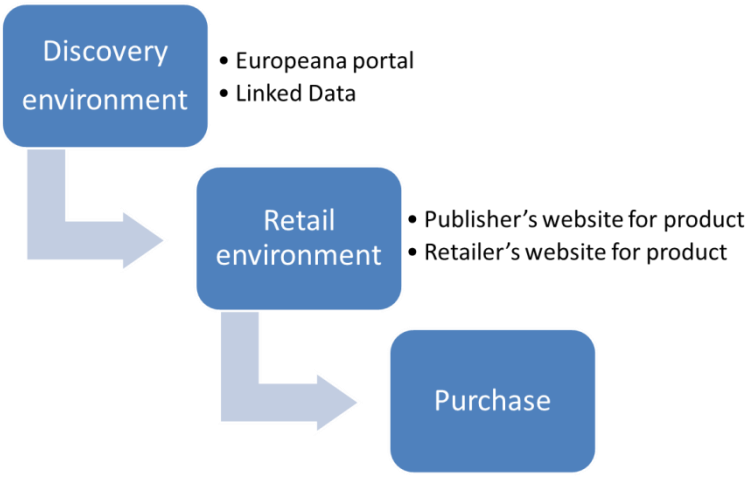


Figure 2 - user / customer experience of Linked Heritage WP4 business model

Although this looks fairly simple, there are outstanding legal and commercial difficulties in the upper two levels of the Figure for the commercial data contributor, and these are discussed in detail in D4.3.

This figure can be reconsidered by replacing 'Retail environment' with 'Cultural heritage institution' and 'Purchase' with 'Licence'. Here the 'customer' may well be a commercial publisher, wishing to license the CHO or DO for reuse. Some distinctive factors in this scenario are discussed here.

3.1 EUROPEANA MATERIALS FOR PURCHASE

"*In Europeana, there are already material*[s] *that can be purchased.*"

This is still the case (in 2013) but four clarifying points must be made:

- Many of the items that are charged for come from *collections digitised by heritage organisations*; these may be *out-of-copyright materials* where a charge is levied to help fund digitisation. Thus they fall outside the scope of Linked Heritage's business model 1 sopra – although they illustrate one important aspect of model 2 sopra.

[3] Vaisberg, L. (2010). FEP response to EU Green Paper A-134-O. Available at: http://ec.europa.eu/culture/our-policy-development/consultation/EU_organisations/A-134-O-Federation_European_Publishers_FEP.pdf

- Many paid-for materials in Europeana are news photo images, contributed *e.g.* through project EuroPhoto[4], which
    - Funds digitisation of photo agencies' archives *i.e.* images that would not, under commercial conditions, have been digitised;
    - Provides public access to the digitised photos under commercial conditions, and specifically, in terms of business-to-business image licensing practice, rather than *e.g.* buying a print of the image[5];

Neither of the two models above fits into the model of providing *both* retail access terms *and* in-copyright content. Of course this does not detract at all from the high *cultural* value of these existing offerings; only that they are not in scope for the current problem, which is purely *commercial*.

Two remaining categories of materials that have been found in Europeana are more illustrative of commercial problems:

- Some of the materials previously available through Europeana against payment have *not* been included in *post-CC0* Europeana aggregations since January 2012:
    - SCRAN (Scottish archive holding "370,000 images, movies and sounds from museums, galleries, archives and the media"[6]). At the time of removal, this single archive represented around 2% of the entire Europeana record set;
    - Ebooks from one large international publisher and one national / language-group book industry data service.

    - The removal of book data is probably due to the intrinsic commercial value of the data and the need to protect the commercial interests or contracted rights of commercial data aggregators and suppliers (see D4.3); in the case of photos, the issue may be the requirement for attribution (and other existing terms of use) *even for thumbnails*[7]. CC0 threatens both of these common commercial frameworks, as discussed in detail in D4.3.

- Metadata records[8] for music recordings from The Orchard music distributor. These are in-commerce, in-copyright products – but because no links to retail pages are supplied in the data, even though a user could perform a new search to obtain the specific track or album found (perhaps with some difficulty), *discovery is separated from access to the full content.*

    - This is a commercial consideration raised by the possibility of *multiple retail offers* for any one product (album or track). Unless Europeana could offer (and maintain) multiple links to commercial outlets for each product record[9], there could be at least the appearance of offering an unfair competitive advantage to one retailer over others. Again, D4.3 details the possible ramifications of this position and offers recommendations.

The latter two examples show that the general conclusions of D4.3 are still valid: under current circumstances, contribution of product data containing retail links is difficult. Solutions could include:

- *Alternative licensing terms* for commercial data & richer previews;

[4] http://www.europhoto.eu.com/Project.aspx - and see also http://www.europeana-photography.eu/index.php?en/57/about-us, another Europeana project with similar business model.
[5] At least one provider, Leodis (with 57,910 records) does provide prints to the general public; by definition these are not highly "commercial" images.
[6] http://www.scran.ac.uk/info/aboutscran.php
[7] See *e.g.* http://www.scran.ac.uk/info/copyright.php and detailed discussion in D4.3
[8] 283,207 records, as of June 2013, but both albums and single tracks are assigned their own records these data so the number of records is somewhat inflated.
[9] Or one resolvable identifier providing multiple links (amounting to the same functionality, but possibly with a third-party provider such as a DOI Registration Agency or European aggregator).

- Managed or outsourced *product identifier resolution services*;

These are also recommendations of D4.3 and can be found in more detail there.

3.2 CLASSIC DIGITAL LIBRARY INTERFACES

> *"the cultural heritage material leads to a commercial front-end, either through a classic digital library interface…"*

This comment accurately describes Linked Heritage's approach, which we consider justified, but perhaps not yet adequately explained and extrapolated. We attempt to improve on our explanation here.

3.2.1 Business models will be limited in the current framework

We fully acknowledge the point made here by the reviewers, that we only considered a very narrow range of business models. In fact, as explained below, we advise that *possible business models will likely be even more limited than they envisaged,* until new technical and licensing developments occur.

3.2.2 There is still scope to expand digital library content

On the other hand, as in 3.1 sopra, Linked Heritage's PPP work showed that there could still be wide scope for increasing the *amount* and *relevance* of commercially available cultural content available across Europe using only the "classic digital library model".

Two of the examples of PPP from D4.1 can be contrasted in to show the potential growth using only this model, when moving from a project to a long-term partnership:

	Gallica 2 BNF (France)	**Enclave BNE (Spain)**
Scope	104,542 records	2,800 records
Model	Partnership	Project
Full-text search index	External	Internal
Preview	External	Internal
Retail links selection	External	External

- Both initiatives took place on a national scale, supported by the national libraries and publishers' associations;
- Both addressed the problem of enabling a selection of retail offers through *external* distributors; this was probably more coordinated in the case of Gallica due to the possibility of offering long-term partnerships to multiple distributors and retailers and thus incentivising them to maintain their links in the hope of making more sales over time; however, Enclave's distributors also offer multiple retailers where possible;
- Both systems offer full-text search (a recommendation of Linked Heritage D4.3) but in Enclave's case this was more difficult because the national library provided this service, as opposed to distributors and retailers in France; text previews could easily be offered internally or externally.
- Of course, the main difference is the huge differential in the number of product records; 50 times more for the "service level" partnership arrangement in France.

The make-up of the publishing and online retail businesses in each country, as well as the public sector technical and financial resources available will play a part in the growth of such services; France's ebook

market is more developed than that in Spain (and part of Enclave's aim was precisely to stimulate growth).

Even so, it would be expected that on a Europe-wide basis, with some central funding, more should be possible, and at least part of the increase will be due to the commitment to a lasting partnership.

3.2.3 A new framework could enable new models

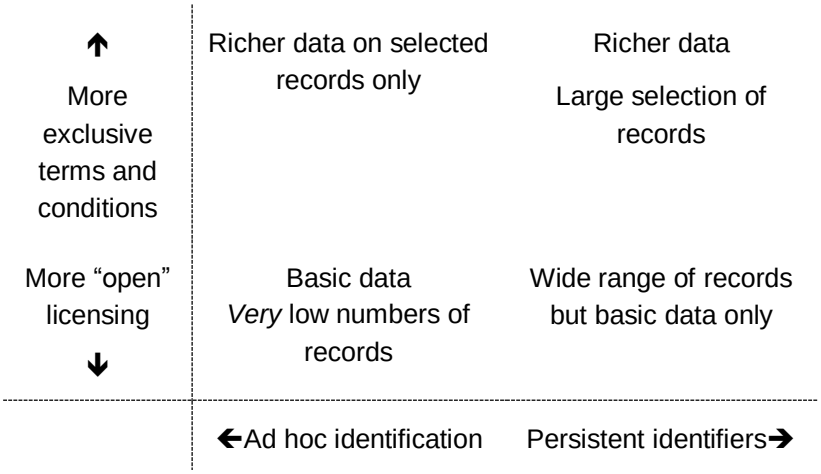
Thirdly, this classic digital library model is imposed by the current business model of Europeana's portal service itself. Two aspects of a new framework could immediately open the door to new business models:

- introducing alternatives to the CC0 DEA for commercial sector data;
- fully exploiting the potential of product (and other content) identifiers.

Europeana is indeed a "classical digital library", and its new DEA has consolidated its basic distinction (almost a separation) of *discovery* from *access* to resources. The latter, *access to resources*, also includes *access to richer information about the resource.*

Here the commercial sector shares a certain tension with the heritage world; as we note again in section 3.3.2 below, ***managed links*** *between different types of digital surrogates have their own value.* Any intermediary which alters the way these links are managed has to either:

- adapt to the – often very specific – ways its partners manage links, or
- compromise on the richness and quantity of data, as summarised in the rough sketch below:



3.2.4 The Centrality of Identification

In the sketch above, the most obvious correlation is between increased control over data reuse terms and conditions, and the richness and quantity of the data that can be contributed.

The less obvious link, represented in the horizontal axis, is between persistent and actionable identification of commercial products and related content and the volume of data records. To put the problem briefly:

Managed systems of persistent identifiers enable ecommerce, and without them, any efforts at public–private partnership with the commercial sector will be extremely limited in scale and value:

- Commercial value is only realised in some sort of supply chain;
- Persistent identifiers enable this supply chain by securing references to products, categories, rights and agreements between partners.

Identification of Cultural Heritage Object and the Digital Objects representing them becomes relevant to commercial partners when it enables:

- licensing the CHOs and DOs to create new commercial products or services;
- adding some value to the "downstream" supply chain by enriching existing products or services *e.g.* through enhancing sales by providing richer discovery environments.

Some areas of research and development in the cultural heritage context point to a new area of common ground which PPP can address in future:

- It is instructive that Europeana is aware of the new issue of *identifiers for cultural heritage objects themselves* in its work on EDM:

> *"Proxies... are introduced to represent the point of view of a specific actor (provider, user, Europeana) on an object.*
>
> *This corresponds to a specific need: the one for distinguishing between different sources of information, when several such sources concern the same object.*
>
> ***Europeana itself would take care of assigning (or re-assigning) URIs for the proxies it creates… It will also create URIs for all objects**, so as to implement a linked data publication strategy that relies on Europeana's own (HTTP) services."*[10]

- The need for identifiers will also arise in attempts to generate cultural value from publishing *Linked Open Data:*

  - Commercial content identification systems have been designed with Linked Data in mind; DOI, in this example, was first widely adopted by commercial publishers (of academic journals) but is increasingly used by research organisations, *e.g.* for raw scientific and social research data, where Linked Data is already available:

> *"A significant advantage of applying Linked Data principles and technologies to DOI-registered material is that it is 'data worth linking to': it is curated, value-added, data, which is managed, corrected, updated and consistently maintained by Registration Agencies.*
>
> *It is also persistent, so avoiding 'bit-rot'. In practice, the quality of Linked data implementations is only as good as the data you are linking to, and the meaning and contextualisation of the link you use.*
>
> *The DOI system can offer "curated data",* i.e. *consistent, managed, linking so you can link to other "quality data" with confidence, while still using the standard Linked Data technologies."*[11]

  - The international heritage information systems community is currently addressing the need to "curate" identifiers as systematically as heritage objects themselves:

> *"When museum objects are referred to in Internet applications, it is necessary that the objects are uniquely identified by suitable URIs…*

[10] See EDM Primer (http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5) at http://pro.europeana.eu/edm-documentation
[11] See http://www.doi.org/doi_handbook/5_Applications.html#5.4

> *The most natural candidate for the URI authority for an object is the museum that curates the object… This is because it is the only institution that can absolutely determine that two different museum object URIs actually describe the same thing"*[12]

> Linked Heritage Work Group 2 has also summarised the need for a more structured approach to identification and identity management in cultural heritage institutions in D2.2.

The three viewpoints cited above represent possible directions that cultural heritage linked data could take. We note that Europeana's remark in the EDM primer appears to point the opposite way to that of CIDOC's recommendation with regard to 'ownership' of the identifier assignment process: CIDOC recommends that authority lies with the curating institution, whereas Europeana suggests it should assign identifiers. On the other hand, some central mechanism to allow a curating institution to "claim" or "verify" URIs minted by Europeana that it recognises as correct representations of the objects it hold, might be able to reconcile these apparently contradictory approaches. A central registry such as a DOI implementation might be one way to achieve this in practice; DOI metadata could mark URIs as verified (or not) by the curating institutions, and also enable handover management of the URI and the associated metadata, should the underlying curated items change hands. This separation of the *identification* from the *current location* (physical or digital) of a cultural heritage object is clearly important it is widely recognised (*e.g.* within such proposals as MuseumID[13] – although the MuseumID URI generation algorithm requires the location, it is cryptographically hashed in the generated URI for the object, and becomes irrelevant in future use of the URI as an identifier. The original location cannot be recovered from the URI.).

[12] See http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html
[13] See http://museumid.net

3.3 LINKED DATA BUSINESS MODELS?

"…or through linked data which can support this kind of business model"

The "classic digital library" business model considered by Linked Heritage has an exact parallel in the model of product discovery and purchase in much of the online retail world, whether for digital or physical products.

3.3.1 Commercial product data is already "linked" data

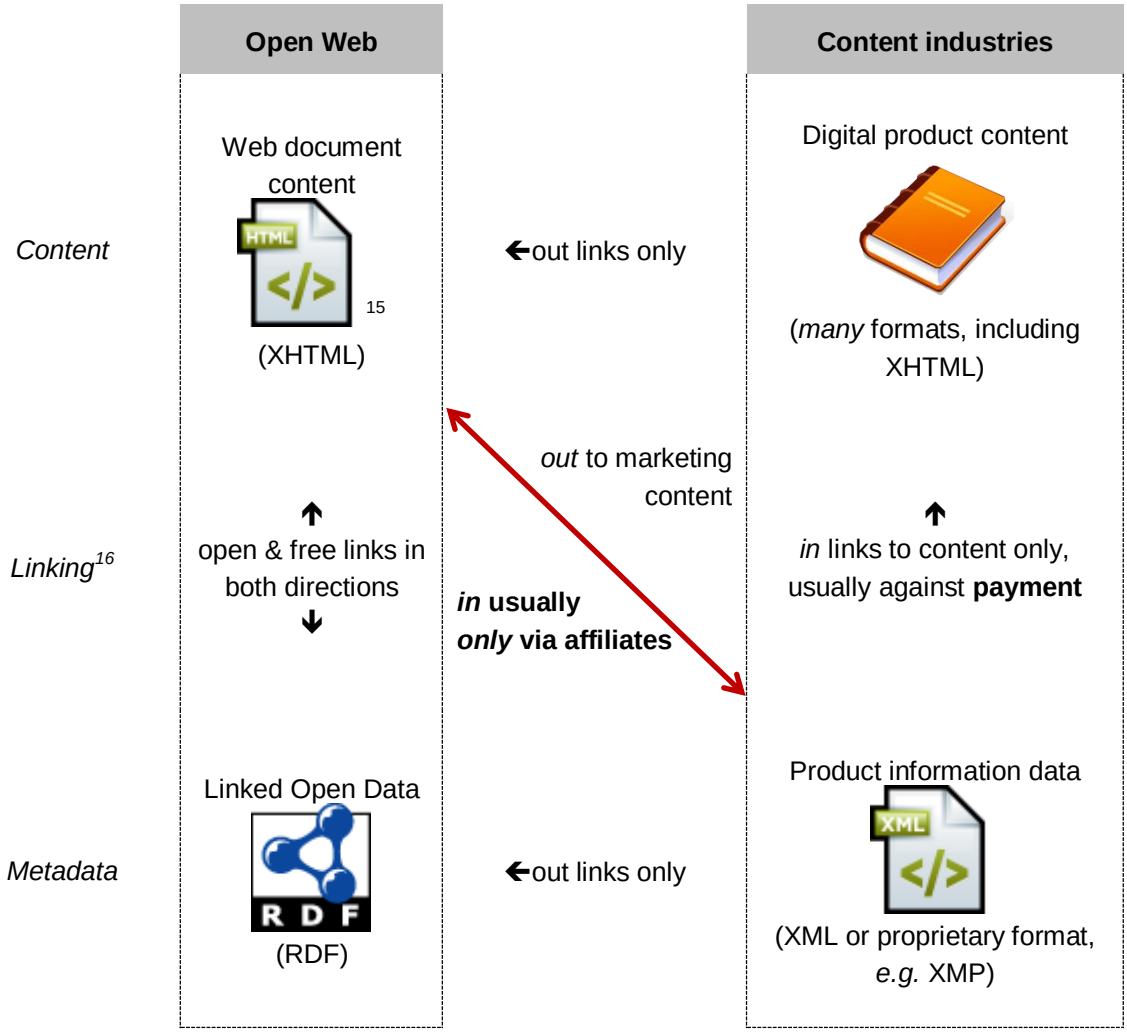
Any contrast between the "classic digital library" and "linked data" must not focus so much on *links* as on *how links are managed.*

In practice, *commercial data is extremely rich in semantic links of many kinds*, primarily through *product, creator and work identifiers* but increasingly through more explicit registration and (actionable) identification of conceptual relationships through controlled value lists (equivalent to "concept schemes" in a popular Linked Data paradigm) such as rich subject or genre classification schemes. These links are not usually expressed as Linked Data (in the purely technological sense), but they create webs of interconnected data entities in much the same way, ideally enabling the data to be traversed by a potential purchaser – from a book to an author, to that same person acting as a musician, to a recorded musical work, to a cover artist… It is "data with links" rather than Linked Data[14].

However, the most relevant contrast is really in terms of control of these links to *guarantee the authority of the data*, and *control of access to data*, both for end users *and* redistributors. The latter directly protects the revenue streams from licensing the data, and the former indirectly protects that revenue by directly protecting the data quality and preventing devaluation of the data through unmanaged reuse. Both aim to ensure its sustainability (as explained in section 3.4.2 sotto and following paragraphs) by enabling the discovery of a content product and an eventual transaction.

The diagram below illustrates the situation for a digital product, for example an ebook, making clear how it is *functionally* like navigation and access to content on the "open Web", but with controlled *access*:

[14] See Bell, G, *Commercial and cultural sectors: potential for data collaboration?,* DOI: 10.4403/jlis.it-5487 JLIS.it (Vol. 4, No.1, 2013) available from http://leo.cilea.it/index.php/jlis/article/view/5487/7867



An analysis of the potential risks of publishing commercial product metadata as Linked Open Data can be found in D4.3. Here we will briefly consider this suggestion in terms of *commercial incentives.*

3.3.2 Linked *open* data could work, but only with a licence

Data supply chains, that produce what a retail customer (or business-to-business content buyer) ultimately finds and uses to make a purchase, normally go in one direction with relatively few reverse flows for a given type of data record.

They may follow a *sequence* of creation, enrichment, combination, and eventual "consumption" (use to make a purchase or other transaction), involving multiple actors in the chain, but almost always, for practicality, in one direction.

Publishing data with broadly open *technical access* (as opposed to a highly permissive or open licence to reuse or modify the data) could potentially enable more reverse data flows and a "re-cycling" of data among partners (primarily this would be better and simpler feedback and correction of the data, the natural corollary of the improved visibility of the data):

[15] HTML and other similar icons from softicons.com - © deleket http://www.softicons.com/free-icons/designers/deleket
[16] In the diagram, "linking" really refers to completed *resolution* of links to content, but is here simplified to give a general impression from a user's perspective.

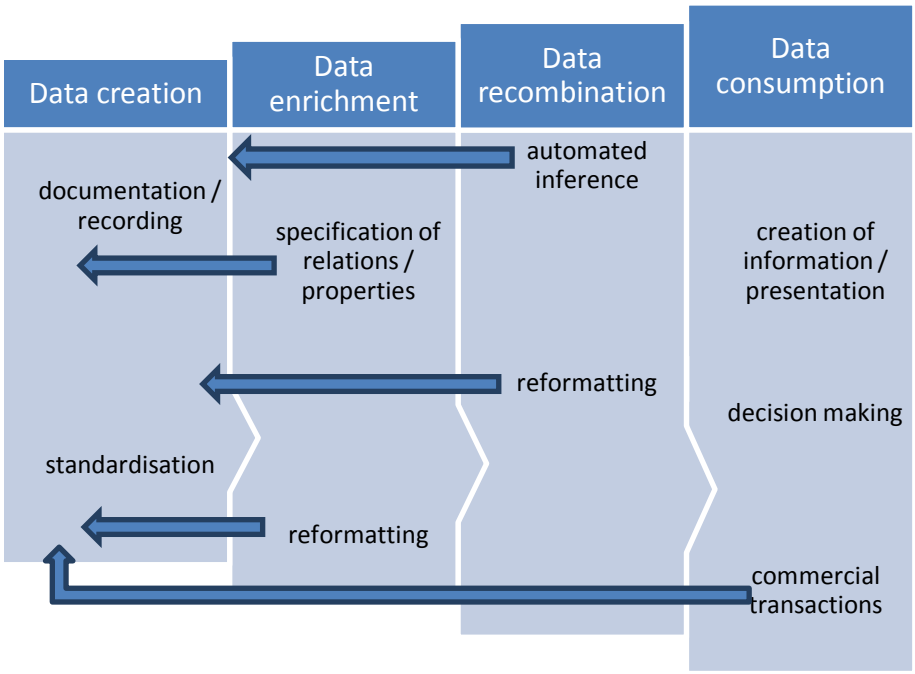


Figure 3 - data supply (left to right) and “recycling” (R to L) in the "open" data paradigm

The diagram in Figure 3 shows the sequence of actions in the data supply chain, broken down into typical activities for each of the four steps noted above. Most of the steps can offer a feedback of enhanced, corrected or otherwise altered data back towards the original provider – and in this sense, they are indeed also found in the normal course of commercial data management.

The lowermost reverse flow – commercial transactions – is the key to maintaining sustainability, by keeping value within this cycle connecting all other users to the data originator.

In principle, openly published data could enhance this process – but only if clear conditions for reusing the data (as in section 3.3.1 sopra) accompany the data publication. If rights in the data are not controlled, then intermediaries in the data supply chain have no incentive to add value – the data originator (e.g. a music or book publisher) and the data consumer (e.g. a music or book retailer) are directly compensated through sales, but any compensation for intermediate management of enrichment of the data would be removed by open data licensing, and the sustainability of the overall business model would be reduced.

This can happen through many effects detailed in D4.3, for which the key source is the “White Paper on Machine-Readable Rights Information” (Bird, 2001)[17].

A CC0 waiver and maximally permissive public licensing *cannot* sustain this type of data cycle; some type of licence ensuring attribution, integrity, ongoing management and other aspects of connection to the data publisher (which are not necessarily incompatible with some other Creative Commons licences) *might* be a starting point for opening commercial data.

[17] Available in archived form at: http://web.archive.org/web/20101003110937/http://www.ldc.upenn.edu/sb/oai/rights.html - and see discussion of CC0 in relation to these issues in Linked Heritage D4.3.

3.3.3 Proposed linked data business models

Several *hypothetical* business models for commercial data to be published in a *technically open* and *permissively licensed* fashion have been proposed:

Source	Proposals	Comments
Open Data Institute[18]	• **freemium**: you provide an "added value" data product or service, for which you charge • **cross subsidy**: you reach more customers, or provide enhanced services to existing customers, through wider sharing and use of your data • **network effects**: by collaborating with other organisations, you reduce your costs in maintaining data which you use in your work or extend the possible audience for your products and services	None of these business models *presupposes* open access or licensing. In addition, "freemium" and "network effects" partially describe the *existing framework* of "linked closed data" in active commerce throughout the media industries.
Scott Brinker[19]	1. **Subsidized/public service**: funded by a government, an NGO, or a regulatory mandate — revenue = funding. 2. **Licensing**: charge fees to let developers use data in other environments. 3. **Microtransactions**: on-demand payments for individual queries or data sets. 4. **Subscriptions**: charge for access to data for a period of time (may have tiers for different levels of access).	Many of these models are in active use with commercial content industry metadata (though this is by and large not 'linked data' in the technological sense, but 'data with links').[20] Some duplicate the ODI's proposals above, with the same *caveats* for commercial

[18] See http://www.theodi.org/guide/how-make-business-case-open-data
[19] See http://chiefmartec.com/2010/03/business-models-for-linked-data-and-web-30/
[20] For example, just one company provides examples of at least seven of these models: Nielsen in the UK provides well-managed aggregated commercial book metadata under strict licence limitations. It lists basic data from publishers for free (somewhat like Freemium, 5.) but charges for inclusion of richer data (Paid inclusion, 6.). It licenses the aggregated data to retailers on a time-limited basis (Licensing, 2. and Subscription, 4.). The data enables an e-commerce service that routes orders from retailers back to publishers, and Nielsen gets a small percentage of each transaction it routes (Affiliate participation, 11). These orders also enable Nielsen to sell monitoring services, essentially top-ten sales charts and market statistics (these are synergistic rather than strictly a Value-add or Loss leader model, 12). Finally, Nielsen has a commercial data-sharing agreement with BDS, a supplier of cataloguing services to the UK national library, so Nielsen indirectly benefits from payments out of the public purse (Subsidized/Public service, 1.).

	5. **Freemium**: free but limited access to data to sample, but charge for extended premium access. 6. **Paid inclusion**: charge to be included in the data set or attributed valuable meta-data (what I formerly called an “authority” business model). 7. **Sponsorships**: charge a small number of advertisers for brand visibility of sponsoring the data. 8. **Advertising**: charge for ads placed around data on web pages (may also tap into ad networks). 9. **Marketplace**: provide data to a partner service in exchange for an opportunistic royalty. 10. **Affiliate program**: provide data streams to affiliates who distribute them in other applications in exchange for commissions on related sales. 11. **Affiliate participation**: as an affiliate of other companies, combine affiliate product links with data to earn commissions on related sales. 12. **Value-add/loss leader**: incorporate free or bonus data as an enhanced feature to win customers for another product or service. 13. **Traffic generation/SEO**: publish data to earn favorable positions in search engines and other directories to generate more traffic. 14. **Branding**: provide data free of charge on a friendly web site to build brand (i.e., self-sponsorship). 15. **Data branding**: provide data free of charge to build brand, but it’s the data itself — not the visible manifestation of it — that is the vehicle for meme distribution.	providers. The remaining two categories of business model (essentially advertising or links to retail sales) depend on the guarantees of integrity and authority described in 3.3.1 sopra.

However, it is worth quoting in full a fully-argued statement on the potential value of linked data using the Traffic generation / Branding business models:

> *“Linked Data mashups (or “meshups” as they sometimes get called) are simply statements linking items in related data sets. Crucially these items are identified by URIs starting “http://”, each of which may have been minted in the domain of the data publisher, meaning that whenever anyone looks up one of these URIs they may be channeled back to the original data source.*

> *It is this feature that creates the business value in Linked Data compared to conventional Web APIs. Rather than releasing data into the cloud untethered and untraceable, Linked Data allows organisations and individuals to expose their data assets in a way that is easily consumed by others, whilst retaining indicators of provenance and a means to capitalise on or otherwise benefit from their commitment to openness. Minting URIs to identify the entities in your data, and linking these to related items in other data sets presents an opportunity to channel traffic back to conventional Web sites when someone looks up those URIs.”*[21]

But in practice, it is *technically very simple* to change the specific URIs that provide ‘indicators of provenance’, allowing abstraction of that traffic and value by a third party, and dilution of any brand value in a ‘commitment to openness’.

[21] From http://tomheath.com/blog/2008/09/where-is-the-business-value-in-linked-data/

3.4 COMMERCIAL DATA AGGREGATIONS

Before moving on from the type of partnership already noted in D4.1 we need to summarise the tensions apparent from the work of Linked Heritage:

- The cultural value of many commercial offerings (data and content) is clear, but a legal and commercial incentive for sharing even minimal metadata widely is difficult to define;
- Experiments with data sharing across cultural and commercial sectors have demonstrated a basic tension whereby the richer, more culturally valuable data will tend to require higher commercial compensation and more restrictive terms for reuse.

We can balance these statements with a few more detailed observations.

3.4.1 Commercial sector partnerships will be culturally valuable

Rich, culturally valuable data about contemporary books, films and TV, music and sound recordings and photographs is already produced in large quantities, in the EU and internationally; in each of these media sectors, standardisation efforts[22] to improve the quality of the data are already mature or rapidly gaining ground, as we noted in D4.1.

D4.2 gave us the chance to estimate how much data could be available overall from each identified "source" in each sector, and how technically viable integrating it might be.

D4.3 presented hands-on case studies of attempts to integrate real datasets from commercial players ranging from niche heritage publishing to international academic presses and dedicated book data suppliers.

The final practical results were:

- A "live", working pipeline for commercial data into Europeana that can be tested with industry standard datasets:
    - Complex data can be fully mapped to LIDO in preparation for later mapping to EDM;
    - Simplified data can be mapped to ESE now with some extra configuration for optimal display;
- A number of identified "target" categories of data contributors for whom the organisation's justification for producing data can align with the case for openly sharing (some) data:
    - Smaller publishers & producers (possibly contacted through technical service providers);
    - Specialised academic publishers;
    - Publishing operations *within* cultural heritage organisations.

Aside from the overall numbers of data records (which in this framework would be small), and the total commercial value of such data, there is at least a highly relevant and relatively manageable starting point here for proofs of concept that can be achieved almost immediately with some additional initiative on Europeana's behalf.

3.4.2 Metadata's value derives from content's value

As was shown in D4.3, the commercial sector's supply chains for metadata depend on the exchange of value based on licensing of intellectual property. Data as such *is* a form of intellectual property (under EU

[22] See Linked Heritage, D4.1. The most prominent standards in each area are:
- Books: http://www.editeur.org/
- Film and TV: http://eidr.org/
- Music: http://www.ddex.net/
- Photography: http://www.iptc.org/site/Photo_Metadata/ and http://www.adobe.com/products/xmp/

law) through the database right, but in commercial terms it is a means to an end (either within a company or as part of a supply chain).

- Ultimately, the presence or absence of the metadata within Europeana has little or no impact on the *intrinsic* value of the CHO or DO for commercial reuse;
- *Discovery* of CHOs and DOs – which Europeana facilitates – is neither difficult nor expensive when compared with the costs of *licensing* those CHOs and DOs for commercial reuse (including discovery of the parties able to licence the CHO or DO, and of negotiating and concluding a licence agreement). As currently structured, Europeana does nothing to facilitate *DO and CHO licensing*.

New projects such as Europeana Creative and Europeana Cloud raise the possibility that precisely this DO licensing can be addressed in future.

On the other hand,

- Even where high quality metadata is created and managed more or less *as a by-product of a business' core activity* (*e.g.* the international academic ebook publisher in D4.3's case study), it is still sound business sense to offer the data as a service for sale:
    - To recoup the costs of maintaining the metadata, or at least the costs of maintaining the metadata supply arrangements (sending a file of metadata may itself be automated and effectively 'cost-free', but a business must also take into account the cost of managing relationships, dealing with data queries that arise, and monitoring the results);
    - To make content deals more attractive by including the relevant metadata as 'added value';
    - To allow the data provider to place licensing requirements upon the data user (such as a requirement for attribution or a service level requirement for updates and maintenance of data quality).

This last point is critical: while Europeana facilitates both the *discovery* of CHOs and DOs (through the portal) and the *dissemination* of metadata (through the API), it does so in an uncontrolled environment that potentially exposes the data provider to a number of risks: reputational risks because of the lack of service level and data quality requirements, branding risks associated with the lack of attribution, as well as direct revenue risks through duplication and cannibalisation of paid-for services.

This means that large-scale commercial data exchanges with Europeana are unlikely to be agreed without a new technical and legal framework constructed in parallel with the existing CC0 structures for cultural data.

3.4.3 Existing cultural content licensing for commerce

A second note of caution must be that

- Many commercial or semi-commercial organisations already manage very substantial numbers of licenses for cultural content, *e.g.*:
    - Commercial picture libraries
    - Educational resources collections
    - Retail products derived from licensed content
- Previous efforts to build large-scale commercial cooperations within the cultural heritage sector alone have encountered significant difficulties, and the full range of lessons learned should be taken into account
    - AMICO and ArtStor
    - RegNet
    - MuseumShop

These clarifications and cautionary reminders should not be taken as a recommendation to simply abandon all efforts at developing public–private partnerships within Europeana, but rather as a realistic warning that sustainable partnerships will require considerable extra time and effort, from both sides, to define and build. More positive indications are noted briefly in 4.2 sotto.

4. ALTERNATIVE TYPES OF PARTNERSHIPS

> *"…investigate different types of partnerships that can be set up and present the potential of such partnerships to various industries…"*

The "different types of partnerships" suggested by the reviewers were:

- *reusing data from the cultural heritage sector within commercial products;*
- *collaboratively publishing data.*

Data reuse was relatively simple to investigate since Europeana is actively promoting it: unfortunately this is where an apparent confusion between *reusing data* (*i.e.* metadata) and *reusing content* in commercial contexts becomes clearest.

The idea of collaboratively publishing data has precedents in past *public–private partnerships for content creation*, as well as the existing *publishing and licensing activities of heritage organisations*.

4.1 REUSING CULTURAL HERITAGE SECTOR DATA

By "reusing data" Europeana often appears to mean *both* reuse of its metadata for API lookups and database enrichment, *and* reuse of the content referred to *by* the metadata. This apparent confusion seems to lead to some difficulties engaging with the commercial sector, which

- Already produces and consumes significant quantities of its own (product) metadata, at cost;
- Also has access to significant services providing metadata on cultural objects, primarily for purposes of picture and music licensing;
- Expects both
    - Clear legal protection of its own metadata and content data, and
    - Certainty and guarantees of the rights status of metadata and content data it reuses and redistributes.

Any new reuse of cultural metadata or content would ideally respect these pre-existing requirements. CC0 would seem to challenge commercial reuse because

- It offers none of the aforementioned guarantees for the *primary reusers*;
    - Clearance of third-party rights embedded within metadata or content is an unknown cost. The CC0 disclaims any guarantees over third-party rights. Licensing of metadata or content for reuse under a normal commercial licence may be less costly, and certainly presents a lower risk, than a diligent search for third-party rights;
- It removes *any* legal obligation for *secondary or further reusers* to pay back any of the extra value created in the first stage of commercial sector reuse.

This concern was raised in D4.1 and D4.2; in the final D4.2 report, we also collected strong statements from leading commercial players confirming it, and discovered that based on the evidence collected, real data *contributions* would be extremely limited under CC0 terms. Data *contributors* are unlikely to undermine the existing data services that commercial partners already show that they value (by paying for the right to reuse them).

For the current work, we attempted to find an evidence base to test these concerns about *reuse*.[23]

[23] Note that such reuse of Europeana metadata is not in reality a 'public–private partnership', since Europeana itself has disclaimed all interest in the metadata when publishing it under a CC0 waiver.

4.1.1 Survey of existing re-uses

In order to assess the reuse potential of Europeana's existing CC0-licensed (ESE) dataset we focussed on the *actual results to date* of attempts to reuse this metadata via the Europeana API[24]. The intention was to investigate how the data has been used *in practice*.

Given that the CC0 licence allows free and unrestricted commercial use *of the metadata itself* two hypotheses about commercial reuses come to the fore:

- Any commercial companies aware of the potential benefits of using this data can do so without charge or legal obligations. So given that it would be commercially attractive, we would expect at least some to "test the waters";
- Further, if any experimental apps showed commercial promise, their code and basic design concepts could be legally protected even if the underlying (CC0) database is not; hence they would be subject to exclusivity, and we might expect at least some pilot apps to be released soon after in commercial versions.

How many and *how commercial* these apps would be was the question posed.

4.1.2 Survey sources and method

The survey proceeded by

- collating already documented apps and API implentations,
- adding new information about their degree of commercial potential,
- and making basic initial categorisations and comparisons.

The starting point for identifying cultural APIs, apps and implementations was of course Europeana's own Web site. More were found through following the work of other researchers contacted through Linked Heritage WP4's work on public–private partnerships, and the wider project context of linked data and identifiers for museum objects:

- The Europeana v2 project produced a "Core Inventory of FLOSS in the Cultural Heritage Domain"[25];
- "Museums and the machine-processable web" wiki maintained by Mia Ridge[26] contains a further, continually-updated list of apps and API implementations;
- Further apps were discovered by following Web links for hackathon events and museum APIs, discovered from straightforward Web searches.

Once these basic facts were documented, *additional value* was added by

- finding the apps themselves, and determining if they were openly available or on sale, still usable (if freely found on the Web) and in what stage of development they were;
- tracking down the developers of each app, where possible, to ascertain some of the above details, and learn more about the developers themselves.

[24] See http://pro.europeana.eu/web/guest/api. The API differs from provision of Linked Open Data in establishing a direct relationship between the provider and reuser of the data, maintained technically through registration and the need to retrieve data through "calls" on the Europeana servers. Linked Open Data publications, once released under CC0 could be redistributed and reused with no obligation to inform Europeana of these activities, or attribute the data to Europeana and its contributors; nor would any identifiable traces have to remain, such as distinctive URI strings for objects and relationships. Neil Wilson, Head of Metadata Services at the British Library, has commented (in personal communication) that while the Library can quantify the load placed on their servers by its Linked Open Data trials – in terms of downloads, API calls, Sparql queries *etc*. – the *net results* of this load are unknowable since the Library has no knowledge of the uses to which the data is put, nor any indication who the users of the data are.

[25] Available here: http://pro.europeana.eu/documents/866067/39f41b8f-f924-4ab5-9394-2e079e43e98c and as a "living document", here: https://docs.google.com/spreadsheet/ccc?key=0Ag_7rVJwt0CpdFRJOEJxdEk4ZEMxQ01jaDgxQXFSTkE

[26] http://museum-api.pbworks.com/

The information was *collated and organised* according to simple criteria, giving an initial indication of their commercial potential:

- cost to obtain or use the app;
- accessibility through retail channels;
- source code availability;
- “owners” – individual developers or companies.

The results were also shared with colleagues in the Linked Heritage project, wider Europeana Network and interested linked data activists.

4.1.3 Survey results

Our simple survey resulted in a data set covering:

- all known Europeana hackathon apps over the period 2011–2012;
- all known implementations of Europeana’s API through a European project or national institution’s Web site;
- all applications listed on the “Museums and the machine-processable web” wiki at the time of the survey (2012);
- other hackathons around the world, including:
    - GovHack 2012;
    - Art Bytes hackathon at the Walters Art Museum;
    - Cosmic Collections.

In total, 99 apps or implementations were documented – none of which have resulted in a significant commercial product. Full results are found in appendix 6 sotto.

4.1.4 Types of application and API implementation

The survey found only two main categories of software (though of course these were assumed by the aims and methodology of the survey, *no new types* were identified):

- API implementations – so called to distinguish these applications from stand-alone “apps”, these integrated the data from the API into an existing application to enhance its search results
- “apps” developed as complete, self-contained units for one of several main functions:
    - To provide a more intuitive *search interface* for cultural data;
    - To *display search results* in a particular way, such as on a timeline or by geographical location of the underlying Heritage Object (this type made up the vast majority of apps);
    - To *enable users to contribute,* by adding data, sharing links via social media, or engaging in some kind of game or shared task, like creating a virtual exhibition.

The seven apps that were available for retail at a price (see 4.1.8 sotto) fell into three specific categories:

- Visualisation of images by some feature of their metadata;
- Random display of images;
- Geographic search by location of the underlying object.

4.1.5 Types of development organisation

All of the applications were developed by one of three types of organisation:

- A cultural heritage institution or project consortium – this holds for all of the “API implementation” types of applications;
- A small company (all apparently less than 10 employees, most less than 5);
- An individual developer and/or heritage information expert.

Almost all of the standalone apps were developed by the last category above, a single individual or a small team of software developers, often including information consultants with pre-existing links to the cultural sector other heritage experts. This is not unusual compared with similar hackathons held in other areas of experimental IT.

4.1.6 Patterns of product development

Investigation of the development status of each app was hampered by the unavailability of comprehensive documentation for all of the hackathon events, and the difficulty of contacting developers (or even lack of their contact details); most did not respond to emails.

Therefore the data on production status below really reflect the lack of information – and an apparent loss of the initial interest in development.

Production status	Number of apps
Unable to discover current status	83
Project suspended	6
In production	1
Open source code	1
Complete and available	8

One tentative conclusion can be drawn from this partial state of the published information; this probably reflects an experimental first trial of the data, aimed at research or early stage 'proofs of concept', rather than a systematic process aiming at commercially viable products. But it is fair to expect that the developers of any successful experiments moving toward commercial exploitation would be contactable.

The "unable to discover" and "project suspended" categories account for all but one of the Europeana API apps, and suggest that a relatively small proportion of hackathon-style applications progress beyond the initial development work.

The apps termed "complete" are those currently available to the consumer (e.g. via a retail app store). The single "in production" app has a quoted price but is not yet available (see 4.1.8 sotto).

4.1.7 Patterns of reuse and recombination

Only four stand-alone apps apparently use more than one API or dataset.

App Name	Data sets	
Rijksmuseum Spotify	Rijksmuseum API	Spotify API
Riiksify	Rijksmuseum API	Spotify API
Jeremy Ottevanger's mashificator	Europeana API	Yahoo!
ManyEyes visualisations of Cooper-Hewitt and NMSI/Science Museum Group data	Cooper-Hewitt	NMSI/Science Museum Group

This could indicate an unwillingness or inability of most developers (so far) to combine heritage data from with data from different sources. It *could* call into question the *suitability for recombination* of the particular data formats investigated, particularly because of the inconsistent use of standardised and widely shared persistent identifiers within available heritage data.

4.1.8 Retail availability of apps

Only a small number of completely developed apps from the sample were found to be for sale (or even available for free through commercial outlets). These were *almost all available at a nominal or low fee*; an exception was the ArtSpace app, a random image retrieval service dependent on rental of the hardware to display the images. ArtSpace has been "in production" since May 2012 but (at the time of writing) remains not yet available, so is apparently an anomaly in this category.

App Name	Data source	Price	Type
Rijksquiz	Rijksmuseum API	0.89 EUR	Visualisation
Rijksmuseum Meesterwerken	Rijksmuseum API	0.89 EUR	
Rijksmuseum Collection	Rijksmuseum API	0.89 EUR	
Rijksmuseum	Rijksmuseum API	0.89 EUR	
ArtSpace	Europeana API	10 EUR / month	Random display
Fornfynd[27]	SOCH (Swedish Open Cultural Heritage)	15 SEK (€1.79)	Geographic search
Fornminnen[28]	SOCH (Swedish Open Cultural Heritage)	28 SEK (~ €3.25)	

Sales data was unavailable for most apps. A conversation with the makers of Fornfynd (Tiny Bird Interactive AB – http://tinybird.com) confirmed that this app achieved "one or two sales per day", which suggests a gross annual revenue of perhaps €400 (after payment of fees to the app store / retailer). None of these apps appears to be a commercial proposition, and the investment in development is clearly not justified by the sales revenue generated. (Of course in all cases, the potential sales revenue was not the driver behind creation of the app – though given this, and the fact that free of charge applications are downloaded from app stores much more frequently than paid-for apps, it suggests that there might be a benefit to making the apps available free of charge)

One of the most notable results is that none of the apps which can be bought rely on the multi-source Europeana data aggregation, but on data sets specific to one national cultural organisation, or even one museum.

These apps also fall into very specific categories; the Rijksmuseum apps visualising the content of that specific museum collection; the two Swedish apps placing heritage objects in their geographical context. These more marketable apps seem to rely on very focused functions and underlying data from a relatively limited context, perhaps reflecting the fact that meaningful data integration is known to be more managable when performed on data with homogeneous or more restricted semantics.

4.2 COLLABORATIVE DATA PUBLISHING

The methodology used in the Yellow Milkmaid white paper to analyse the advantages of adopting an open policy in the dissemination of Europeana metadata can be used, with some adaptation, to assess pros and cons for private companies (content providers and/or intermediaries) contributing to Europeana.

In the white paper, the authors list ten "potential benefits of open metadata" that could apply also to providing metadata to Europeana, regardless of the reuse licence terms:

- Increasing relevance (#1);

[27] Available at https://itunes.apple.com/se/app/fornfynd/id443389155
[28] Available at https://itunes.apple.com/se/app/fornminnen/id582327948

- Increasing channels to end users (#2);
- Brand value (#4);
- Desired spill-over effects (#10), and in particular;
- Discoverability (#6) and;
- New customers (#7).

All could be good reasons for commercial operators to provide data to Europeana; however, given the lack of awareness of Europeana (and of linked data, open or not) among publishers, and its position as a heritage education and research portal, rather than a retailer or consumer destination website, there will be some extremely high barriers to convincing individual publishers to commit resources in this area.

Data enrichment (#3) and Building expertise (#9) are specific to the open metadata license, and Specific funding opportunities (#5) and Public mission (#8) are peculiar to the public sector, so these will not be considered.

The current value of Europeana for brand support (#4) is considered low. One way to place a rough indicative value on this is to consider how much advertising revenue would be worth – effectively treating a view of commercially-contributed metadata describing a product as an advertisement for the product itself. Brand advertising, via banner ads on web pages, varies hugely in value, depending among other factors on the audience attracted by the website. A typical figure might be between €10 and €20 CPM (cost per mille). Highly targeted ads displayed to a desirable demographic group would potentially be more valuable, but these figures are already well above average for the web advertising industry as a whole. Given Europeana's published visit and object page view statistics for 2012 (that is, about 8m object page views spread across 2.5m substantive visits), a total CPM of €20 on every object page would have generated gross revenue of approximately €160,000. Note that each object in Europeana gets viewed much less than once per year on average, and each object view might be worth around 2¢. A content creator contributing 1000 records to Europeana might expect to generate less than €10 of gross advertising revenue annually (assuming commercial objects are neither more nor less likely to be viewed than heritage-sector objects).[29]

This illustrates clearly that the only *realistic and significant* value to a commercial content creator arises from direct click-throughs that result in sales (or licensing) of the DO or CHO. Even there, with every object being viewed (on average) less than once per year, the opportunities that *could potentially* result in a click through and a sale are rare.

Note that the value to a commercial organisation of Europeana's ability to disseminate the metadata more widely – to redistribute the metadata to parties the commercial organisation could not itself reach (possibly #10) – is also likely to be very low. The commercial book, music, image and film/TV sectors already support dissemination of the data in various ways, using semantically rich metadata schemes that are tailored to the data models and use cases of each sector (this was widely discussed in D4.1). While this metadata is often available for payment, these payments are rarely a barrier to commercial exploitation of the metadata (or of the products it describes). Dissemination of relatively basic data via Europeana adds little to this. The current inability of Europeana's data infrastructure to provide adequately for the dynamism and change inherent in commercial data (as discussed in the main D4.1 report) also emphasises the value provided by existing commercially-focused metadata schemes and data exchange arrangements.

The others (#1, #2, #6, #7, possibly #10) all amount in the end to increasing product visibility leading to sales or licensing opportunities.

Combining this simple statement with the value chain analysis, we discover that costs and benefits are distributed in an unequal manner between the different actors operating in the market. Publishers potentially

[29] This assumes also that the metadata can be kept up to date. An object view in Europeana that leads to increased interest in a book that is now our of print, or in a commercial recording on CD that has been deleted from the catalogue is clearly 'worth' much less. The dynamic nature of the commercial data (and Europeana's current inability to track such dynamic data) is discussed further in D4.1 and D4.3.

enjoy the largest benefits (from sales and licensing), but also have by far the highest costs. Beside this, publishers – on average across many European countries – have metadata of lower quality than bibliographic agencies and Books in Print services. If they provide to Europeana directly, we might expect the problem of "yellow metadata". The quality of metadata may be rich enough, but accuracy may be poor, and over time, the reduction in authority and trust compromises the sales and promotion benefits expected from Europeana. Bibliographic agencies and Books in Print services have the lowest costs (since they already act as aggregators and deal with the metadata in bulk) and they have the most accurate metadata (they often add editorial value to the metadata, and they combine information from multiple sources – publishers, distributors and wholesalers – intelligently), but they enjoy the benefit only in highly indirect ways. The other intermediaries have no right to pass on the metadata, since they acquire them from publishers or bibliographic agencies with limited usage rights.

It would be possible to follow the same model adopted for the public institutions: to have a metadata aggregator that intermediates between individual publishers and Europeana. However, since at present, they do not participate in sales, this entails re-organising bibliographic and other data aggregation agencies and redefining how they are compensated to ensure they benefit from any increased sales, which could be very expensive and would possibly enter into competition with existing companies.

Thus the inclusion of large quantities of commercial metadata into Europeana is most likely a cooperation between publishers, who have the interest, and bibliographic or other aggregation agencies, who have the expertise. Only such cooperative efforts may solve the identified trade-off.

The most significant example of a source of commercial book metadata that has joined Europeana is Libreka.de, and it seems to confirm this conclusion[30]. The initiative comes from the publishers' community since MVB – the company that has launched the Libreka e-book venture – is jointly owned by the German publishers and booksellers association. MVB is the reference bibliographic agency and books in print service within the German linguistic area, and Libreka is at the same time an aggregator, a distributor and a retailer of e-books. They combine the ability to deliver metadata in bulk with the ability to benefit financially from any extra sales that metadata generates. However it is also difficult to replicate the same experience in other countries where the circumstances are not the same.

To a great degree, the reuse of Europeana metadata and links to enrich existing commercial-sector metadata could follow the same model too, though in this case the most likely partnership would be between the data aggregation agencies (who contribute expertise) and retailers who benefit directly from the enhancements and deep cultural links that can be added to the commercial metadata and which can enrich the retail experience. Retailers who pay aggregators for commercial data in the books sector, for example, have in the past paid premiums for the integration of independent third-party datasets such as *Kirkus Reviews*[31] into the metadata provided by publishers, and they continue to pay for the 'editorial' or metadata curation work provided by the aggregators. Given the acknowledgement that enriched metadata generates incremental sales[32], such partnerships are not impossible – though because of the inconsistent use of public, persistent identifiers within some of the cultural heritage metadata that is available, and the use of *different* and non-interoperable identification schemes from those used within the commercial data in other cultural heritage metadata, reliable and meaningful enhancement of the metadata with extra links and context drawn from Europeana remains challenging.

[30] However, only experimental metadata has been delivered by MVB so far.
[31] *Kirkus* was formerly owned by Nielsen, which has a significant business unit aggregating publishers' metadata and providing it to the book trade in a reliable, stanardised format
[32] See http://www.isbn.nielsenbook.co.uk/uploads/3971_Nielsen_Metadata_white_paper_A4(3).pdf

5. CONCLUSIONS

5.1 RESULTS FROM SURVEY OF APPS REUSING CULTURAL HERITAGE DATA

The survey described in Section 4.1 illustrates the difficulty of taking existing cultural metadata and extracting commercial value. None of the apps surveyed has turned from ‘interesting experiment’ into a significant revenue source. None of the small number of apps that have become commercially available used cultural metadata to enhance existing private metadata.

Too few app developers could be contacted to draw conclusions about why this is so, however, the following possibilities are likely to be among the key reasons:

- the low awareness of Europeana among commercial organizations who might wish to reuse or provide access to Europeana metadata in their own products
- technical difficulties due to inconsistent identifier policies (and particularly the lack of standardized persistent identifiers) in the cultural sector, and lack of compatibility between commercial and public-sector identifier and metadata schemes (Europeana metadata is not available in the metadata schemas most commonly used by commercial organisations);
- the lack of any exclusivity in the reuse of metadata under CC0 terms means that any development may be ‘undercut’ by other parties;
- availability of lightweight metadata (whether under CC0, under a permissive licence, or on more restrictive terms) is not the important issue. The real value lies in the commercial exploitation of the underlying digital representations of cultural objects. *It is the licensing of the DOs and CHOs themselves that is the key*. Future Europeana-related projects should focus on this issue.

5.2 LIKELY SOURCES OF BULK METADATA FROM THE COMMERCIAL SECTOR

The discussion of likely models for collaboration in data delivery between private sector content creators and metadata providers, and Europeana in Section 4.2 concludes first that the value of such collaboration to the commercial partner is likely to be low, and second that the most desirable partners (from the point of view of data quality, richness, and efficiency) are perhaps the least likely to collaborate:

- the value of exposure and potential redistribution of the metadata is limited by
    - relatively low usage of Europeana among consumers, students, educators and researchers
    - the fact that the Europeana metadata model presents much less rich data to the Europeana user (or to the reuser of the Europeana metadata) than do the various industry-specific metadata schemas;
- the risks of exposure of metadata under CC0 terms is significant (primarily the risk of making existing and valuable data management practices unsustainable, but also reputational risks associated with the lack of attribution, clear provenance and service level control, as well as direct revenue risks through duplication and cannibalisation of paid-for services);
- any benefit of extra sales of a commercial product accrues to the retailer and the content publisher. The most likely source of high-quality metadata – various industry aggregators such as Books in Print services – would not benefit. Cooperation arrangements between publishers and a bibliographic aggregator would need to be set up.

6. APPENDIX – SURVEY SAMPLE

This table lists the applications include in the survey of applications based on Europeana's and other cultural heritage APIs and open datasets. Only the apps and implementations are named, the events or projects from which they originated (if given) and the datasets upon which they were based.

Other details collected include personal data such as names and contact details, development status and availability, prices, etc., and are not exhaustively listed here. Where relevant, these data are given in the body of the report.

Name of app	Event / project originating app	Dataset
365 Days of Art – The virtual Rijksmuseum		Rijksmuseum API
Arkyves ICONCLASS Browser		Rijksmuseum API
Art Galleries for Mobile		Rijksmuseum API
Avtryck		SOCH (Swedish Open Cultural Heritage)
Biblioteca Virtual Ignacio Larramendi		Europeana API
ConnectedCollection		Rijksmuseum API
Digital Humanities Observatory		Europeana API
Doekvoorinjehoek.nl		Rijksmuseum API
Dutch Masters Mobile Print Solutions		Rijksmuseum API
Dutch Museum of National History		Europeana API
Europeana Geo Search app for Android		Europeana API
europeana4j		Europeana API
Faces of the Rijksmuseum		Rijksmuseum API

Name of app	Event / project originating app	Dataset
Finnish Digital Library		Europeana API
Fornfynd		SOCH (Swedish Open Cultural Heritage)
Fornminnen		SOCH (Swedish Open Cultural Heritage)
freeuse.copyright.or.kr		Europeana API
Historiska museet		SOCH (Swedish Open Cultural Heritage)
Inventing Europe		Europeana API
Jeremy Ottevanger’s mashificator		Europeana API, Yahoo!, Google, Zemanta
Kringla (also for Android)		SOCH (Swedish Open Cultural Heritage)
Locago		SOCH (Swedish Open Cultural Heritage)
ManyEyes visualisations of Cooper-Hewitt and NMSI/Science Museum Group data		Cooper-Hewitt, NMSI/Science Museum Group
National Library of Ireland catalogue widget		Europeana API
Open Archive Twitter Streams		Rijksmuseum API
Open Images Mash-up		Rijksmuseum API
Paint Job		Rijksmuseum API
Platsr		SOCH (Swedish Open Cultural Heritage)
Riiksify		Rijksmuseum API, Spotify API
Rijksmuseum		Rijksmuseum API

Name of app	Event / project originating app	Dataset
Rijksmuseum Collection		Rijksmuseum API
Rijksmuseum Meesterwerken		Rijksmuseum API
Rijksmuseum Quiz		Rijksmuseum API
Rijksmuseum Spotify		Rijksmuseum API, Spotify API
Rijksmuseum-objectentijdlijn		Rijksmuseum API
Rijksquiz		Rijksmuseum API
Royal Museum for Central Africa search		Europeana API
Statenvertaling.net - The Bible and Art		Rijksmuseum API
Tekniska museet		SOCH (Swedish Open Cultural Heritage)
Västarvet		SOCH (Swedish Open Cultural Heritage)
VuFind Europeana search results recommendation module		Europeana API
What is this?		CSV-downloadable collections of the Science Museum Group
Europeana API geo-wrapper	1st Europeana Hackathon, 1-2 April 2011, Hilversum, Netherlands	Europeana API
Europeana Banana	1st Europeana Hackathon, 1-2 April 2011, Hilversum, Netherlands	Europeana API
Images explorer	1st Europeana Hackathon, 1-2 April 2011, Hilversum, Netherlands	Europeana API

Name of app	Event / project originating app	Dataset
Wikimedia batch upload	1st Europeana Hackathon, 1-2 April 2011, Hilversum, Netherlands	Europeana API
Badgify the Walters	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Dave Raynes	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Frame	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Pez-Head	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Schroddon	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Tanzaku	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
The WalTours	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
Time Machine	Art Bytes hackathon at the Walters Art Museum	Walters Art Museum API
ATHENA	ATHENA	Europeana API
ECLAP	ECLAP	Europeana API
Europeana eCloud	Europeana 1914-18	Europeana API
Europeana Remix	Europeana 1914-18	Europeana API
DISMARC	Europeana Connect	Europeana API
Europeana Local Austria	Europeana Local	Europeana API
Culture Globe	EuropeanaTech hackaton, 3-5 October 2011, Vienna, Austria	Europeana API

Name of app	Event / project originating app	Dataset
Image form Query suggestions	EuropeanaTech hackaton, 3-5 October 2011, Vienna, Austria	Europeana API
Storyana	EuropeanaTech hackaton, 3-5 October 2011, Vienna, Austria	Europeana API
A Day in the Life	GovHack 2012	GovHack 2012 archives dataset
History in ACTION	GovHack 2012	GovHack 2012 archives dataset
Photo Search	GovHack 2012	GovHack 2012 archives dataset
Europeana4Education (E4E)	Hack4Europe! Belgium, 13-15 June 2012, Leuven	Europeana API
Portal Thingy	Hack4Europe! Belgium, 13-15 June 2012, Leuven	Europeana API
Stackathon	Hack4Europe! Belgium, 13-15 June 2012, Leuven	Europeana API
Storyana	Hack4Europe! Belgium, 13-15 June 2012, Leuven	Europeana API
europ.in	Hack4Europe! Latvia, 26-27 May 2012, Riga	Europeana API
History4you	Hack4Europe! Latvia, 26-27 May 2012, Riga	Europeana API
Postcards from Baltics	Hack4Europe! Latvia, 26-27 May 2012, Riga	Europeana API
The Culture Center (TCC)	Hack4Europe! Latvia, 26-27 May 2012, Riga	Europeana API
Art Seeker	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API
Art Terra	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API
ArtSpace	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API

Name of app	Event / project originating app	Dataset
Institution recommendations based on geospatial data	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API
Photomaze	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API
Virtual polymath - A multidimensional approach to art and culture	Hack4Europe! Poland, 26-27 May 2012, Warsaw	Europeana API
Art4Europe	Hack4Europe! Poland, 7-8 June 2011, Poznan	Europeana API
Europeana Field Game	Hack4Europe! Poland, 7-8 June 2011, Poznan	Europeana API
HackMemory	Hack4Europe! Poland, 7-8 June 2011, Poznan	Europeana API
Art Colour Bits	Hack4Europe! Spain, 8-9 June 2011, Barcelona	Europeana API
Europeana Layer	Hack4Europe! Spain, 8-9 June 2011, Barcelona	Europeana API
TimeBook	Hack4Europe! Spain, 8-9 June 2011, Barcelona	Europeana API
Visual Chronology	Hack4Europe! Spain, 8-9 June 2011, Barcelona	Europeana API
Europeana Fancy Search	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
Europeana on Free Image Search Tool	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
Europeana Portfolio	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
MuseumVille	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
myHometo.wn	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
Time Mash	Hack4Europe! Sweden, 10-11 June 2011, Stockholm	Europeana API
Casual Curator	Hack4Europe! UK, 9 June 2011, London	Europeana API

Name of app	Event / project originating app	Dataset
Cultured Canvas	Hack4Europe! UK, 9 June 2011, London	Europeana API
euKindleSearch	Hack4Europe! UK, 9 June 2011, London	Europeana API
Share What You See	Hack4Europe! UK, 9 June 2011, London	Europeana API
HISPANA	HISPANA	Europeana API
PIONIER	PIONIER	Europeana API