DELIVERABLE

Project Acronym:	**LINKED HERITAGE**
Grant Agreement number:	**270905**
Project Title:	**Coordination of standard and technologies for the enrichment of Europeana**

D4.3 Specification of legal/licensing environment

Revision: version 1.0

Authors:

Michael Hopwood (EDItEUR)

Contributors:

Piero Attanasio, Paola Mazzucchi and Vincenzo Cesino (mEDRA)

Reviewers:

Dimitrios K. Tsolis (UP)

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

Revision History

Rev	Date	Author	Organisation	Description
0.1	1/3/2013	Michael Hopwood	EDItEUR	First rough draft
0.2	7/4/2013	Michael Hopwood	EDItEUR	For internal (EDItEUR) review
0.3	22/5/2013	Michael Hopwood	EDItEUR	For Internal (project WG and LH project) review
0.4	7/6/2013	Michael Hopwood	EDItEUR	For submission to Commission
1.0	13/06/2013	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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Contents

EXECUTIVE SUMMARY 6

1 INTRODUCTION 11

1.1 Background to this Deliverable 12
1.2 Aims of this Deliverable 13
1.2.1 Task T4.2 – Contribution specification 13
1.2.2 Other Work Package Objectives 15
1.2.3 Wider Aims of Linked Heritage 15
1.3 Methods Employed to Produce this Deliverable 15
1.3.1 Literature Review 16
1.3.2 Best Practice Case Studies 16
1.3.3 Expert Interviews leading to Business Model Case Studies 16
1.3.4 Development of Cooperation Agreement for Commercial Sector ("Prototype") Data Contributors 16
1.3.5 Informal survey of attitudes to contribution and commercial reuse 17

2 LEGAL FRAMEWORK – COPYRIGHT AND DATABASE RIGHT 18

2.1 Legal Foundations – What Constitutes Content? 18
2.1.1 Copyright in Creative Content 22
2.1.2 Are Items of Metadata Under Copyright? 22
2.1.3 Databases are Covered by Database Right 25
2.2 Legal Framework – conclusions 26

3 LICENSING FRAMEWORK 27

3.1 Implied Licences 28
3.2 Specific End-User Licences 30
3.3 Creative Commons Licences 33
3.3.1 CC Licences: Reproduction, Distribution, Attribution, Derivation, Commercial Use, Like Terms, and Licence Notices 33
3.3.2 CC0 1.0 Universal (CC0 1.0) - Public Domain Dedication 33
3.3.3 Potential issues with fully open access metadata sets 35
3.3.4 Linked Heritage Response to the Europeana DEA 36
3.3.5 Current Linked Heritage Aggregator Configuration 37
3.3.6 Types of content and rights covered by Europeana's DEA 38
3.3.7 Current Europeana Licensing Framework 40
3.4 Licensing framework - conclusions 41
3.5 Licensing frameworks – metadata or content? 41

4 BUSINESS MODELS AND SUSTAINABILITY 44

4.1 Quid quo Pro – retail links for product metadata 44
4.1.1 Cost and benefits beyond the basic exchange model? 44
4.1.2 The content and metadata value chains in European e-book markets 45
4.2 Technical Requirements – Linked Heritage and Europeana 49
4.2.1 Europeana mandatory elements 49
4.2.2 LIDO mandatory elements 53
4.2.3 MINT requirements 54
4.2.4 Minimal requirements of commercial sector data providers 54
4.3 Prototype Datasets within Linked Heritage's Lifecycle 56
4.3.1 Prototype agreement 57

4.3.2 Paraphrase of the prototype data contribution agreement 57
4.3.3 Terms and conditions 58
4.3.4 The missing link – agreements with Europeana 58
4.3.5 Data owners and data service providers 60
4.4 Dimensions of Business Models 61
4.4.1 Business Models for Content Aggregation – "the partnership" 61
4.4.2 Defining the partnership – contracting partners 62
4.4.3 Commercial sector product data – "the offering" 63
4.4.4 Defining the offering – whose product? 64
4.4.5 Data as a core business asset – "the service" 64
4.4.6 Defining the service – exposure 65
4.4.7 Defining the service – coverage 65
4.4.8 Defining the service – currency 66
4.4.9 Quantity of data contribution – "the dataset" 67
4.4.10 Defining the dataset – Europeana's stated aims 67
4.4.11 Defining the dataset – product form 71
4.4.12 Defining the dataset – subject classifications 71
4.4.13 Quality of data contribution – "the template" 72
4.4.14 Defining the template – included information 73
4.4.15 Defining the template – granularity of data 73
4.5 Case Studies – Commercial offerings & degrees of openness 74
4.5.1 Heritage E-books (image library and associated e-book publisher) 75
4.5.2 Large image library 78
4.5.3 Large academic library's publishing arm 81
4.5.4 National library's publishing arm 83
4.5.5 Large academic e-book retailer 85
4.5.6 Large national academic e-book publishers 88
4.5.7 Large international academic e-book publisher 92
4.5.8 Medium sized book data service providers 94
4.5.9 Large national book data service providers 95
4.5.10 Large international book data service providers 105
4.6 Sustainability 107
4.6.1 Continuity of Management – which partners? 108
4.6.2 Management by public body 108
4.6.3 Management by commercial organisation 108
4.6.4 Data ownership and licensing 109

5 POTENTIAL INCENTIVES FOR COMMERCIAL DATA PROVIDERS? 110

5.1.1 Existing business models of bibliographic agencies 111
5.1.2 Incentives for contributing to Europeana for bibliographic agencies? 113
5.1.3 The use of ISBN-A to provide multiple links via a single URI? 114

6 POTENTIAL REVENUE STREAMS 116

6.1 Public funding 116
6.2 Private funding 116
6.3 Alternative models reducing cost / revenue streams 117

7 POTENTIAL COSTS 119

8 CONCLUSIONS AND RECOMMENDATIONS 120

8.1 for Europeana 120

8.1.1 Dedicate new resources exclusively to acquiring and maintaining commercial data 120
8.1.2 Use separate licensing frameworks for basic and rich metadata 121
8.1.3 Support multiple resolution of unique product identifiers 123
8.1.4 Enable separate search of textual "preview" material 123
8.1.5 Support EC funding for set-up of aggregators or data adaptors for in-copyright materials. 124
8.2 FOR THE LINKED HERITAGE CONSORTIUM 124

9 GLOSSARY OF TERMS 125

10 REFERENCES CITED IN THIS REPORT 129

11 APPENDIX – BOOKS SUBJECT CLASSIFICATION HEADINGS FOR DEFINITION OF A DATASET TO CONTRIBUTE TO EUROPEANA 133

11.1 BISAC 133
11.2 BIC 157
11.3 CCE 184
11.4 WARENGRUPPEN 2.0 191
11.5 DEWEY DECIMAL 193

12 APPENDIX – TEXT OF NEW COOPERATION AGREEMENT FOR COMMERCIAL SECTOR ("PROTOTYPE") DATA CONTRIBUTORS 198

13 APPENDIX – RECIPE FOR CREATING BASIC ONIX 3.0 RECORDS FROM SPREADSHEET DATA 212

EXECUTIVE SUMMARY

This third, final report of Linked Heritage's work on public-private partnership (PPP) seeks detailed business models and a mutually acceptable data licensing framework for contributing commercial product data to Europeana. The recent history of Linked Heritage's engagement with Europeana provides essential context for the concepts and research presented here:

- June 2010: the Federation of European publishers (FEP) stated: "*Links through Europeana to sites of publishers or other rightsholders to enable users to buy in-copyright content are the ideal way to achieve a private sector involvement*";
- September 2011: Europeana announces its new DEA based on CC0;
- December 2011: Linked Heritage WP4 delivers D4.1 *Best Practice Report – Public-Private Partnership* raising some questions as to the viability of any commercial contributions under CC0;
- February 2012: Linked Heritage implements a technical solution to filter its partners' heritage data, allowing publication in Europeana of a subset of a material in which rights have been waived for release under CC0 – as part of this response WP4 develops a distinction between "Test" and "Prototype Data" for commercial data in the project;
- September 2012: the EC requested WP4 to examine commercial reuse of Europeana's data – this will be published as *D4.1 – Addendum*, a miniature research piece comparing business models for linked open data with existing commercial data flows, assessing cultural heritage data-driven apps, and scanning the horizon of cultural enterprises;
- October 2012 – March 2013: having verified technical feasibility, WP4 engaged in conversations with commercial sector content creators and data service providers to describe the commercial considerations, data licensing models and value chain participants informing both this report and the Addendum to D4.1.

The timeline above shows clearly that the Linked Heritage project's PPP work was conceived, planned and begun before Europeana introduced its new data exchange agreement requiring metadata contributions under CC0 rights waiver terms.

There is no doubt that this decision has had an effect on this part of the work of the project, making it more difficult to engage commercial entities and persuade them to contribute their data. The lively commercial market for rich metadata – particularly in book publishing – in many European countries makes delivery of rich data with all rights waived less likely.

However, there are still good reasons to expect this may affect very small publishers (and the publishing arms of heritage bodies) less than large, highly commercial publishers, and this means that there is still a hope that at least some in-commerce products can be represented in Europeana.

As a foundation for discussing business models, this report summarises the legal status of commercial metadata elements, records and aggregations, finding that these always have legal aspects:

- Collections of metadata imply at least a *sui generis database right* in Europe;
- Rich, culturally interesting metadata normally contain some *significant copyright content*.

Although sharing cultural data's focus on description and access, commercial metadata aim to sell products and support other commercial activities linked to sales, a third legal aspect is introduced:

- Any significantly rich or large-scale dataset is really a *data service*, which relies on paid-for data licensing to recoup its costs, ensure the continued viability of the service and, in many cases, generate some profits for the company that provides it.

Thus use of these data in any new context will require explicit, detailed licence agreements, taking into account existing parties' rights and contractual terms.

Value added and received by each partner in the data supply chain was analysed, both in general, and with respect to each data element communicated to Linked Heritage and to Europeana, resulting in a research instrument[1] listing the "dimensions" of data supply business models; this was used to conduct semi-formal interviews and follow-up discussions with commercial publishers and intermediaries ranging from the smallest publishers to the largest international commercial data services.

Two further products of these conversations are presented here:

- The "Prototype Data" framework developed in 2012 was extended with a formal agreement to make it more secure. It was then used as a starting point in approaching small, medium and large book publishers and data services, and data services in other media sectors. *Each term of this proposed agreement represents a minimum licensing requirement captured during the conversations with commercial players;*
- As a key to operationalising the research instrument, a detailed content selection strategy for defining a relevant subset of our contacts' data (where appropriate) was developed and refined in each round of discussions; Europeana's stated content priorities were mapped to the most common commercial subject schemes, and refined against actual commercial product catalogues supplied by publishers.

Responses varied radically in their openness to signing the proposed agreement. No commercial contacts were undecided, but some showed reservation in contributing a full snapshot of their databases, and many declined entirely:

- Unreserved "yes": smaller publishers; publishers based within heritage bodies; "registration" data services;
- Reserved "yes": data services attached to national publishing trade bodies; publisher/distributors;
- "No": purely commercial data services; large e-book publishers; medium image library.

In summary, the commercial players willing to contribute data under the present conditions are precisely those least likely to have high-quality, rich data, in significant volume, in a standardized format – so while, for them, there may be no 'in principle' issues, there are practical issues that prevent large-scale contributions to Europeana.

Based on the evidence of in-person conversations and extensive correspondence with the organisations in the case studies, and many others interested in the project, the key barriers to commercial engagement with Europeana are:

- Lack of awareness of Europeana itself;
- Lack of support within the Linked Heritage and Europeana infrastructure for data supply chain best practice (principally updates for dynamic data);
- Lack of clear commercial incentives, financial or otherwise, that would justify giving up the rights held in the metadata, in terms of
    - Direct payment for provision of data (for aggregators);
    - Potential for increased sales of content items (for publishers or retailers);

[1] Here in D4.3 this was applied exclusively to Linked Heritage's Prototype Data proposal, though in the D4.1 Addendum, it was applied to other forms of Linked Open Data scenarios.

    - Opportunities for enrichment of data to enhance product discovery (for retailers);
- Significant costs to transform and maintain data and ongoing updates to Europeana;
- Threats to existing revenue streams derived from licensing data sets from which Europeana's subset would be taken.

Other commercial business models already in operation in parallel to Europeana's were briefly examined to provide context for the final recommendations of WP4. These recommendations are based on:

- reflection on the size and complexity of the publication subsets produced for the case studies, and datasets of potential not ready to contribute under the existing conditions;
- the experience of WG4 producing test and prototype data in collaboration with commercial contacts, within Linked Heritage;
- approximate numbers of staff and management models of known existing data services, including the Linked Heritage consortium's aggregation platform, the examples of cross-sector data sharing found in Linked Heritage D4.1, and commercial data aggregators interviewed for the case studies.

Ranked by the cumulative level of investment by Europeana (and related projects and organisations) required, and potential benefits for Europeana, they provide an illustrative first step towards a detailed business options analysis that could inform future projects:

Level of investment (cumulative)	Commercial contributor	Management model(s)	Envisaged enrichment of Europeana
Low **(staff with relevant standards and business expertise)**	Small publishers (via data service providers); heritage bodies' own commercial content production (library publishers, sound and film archive re-releases)	Staff to manage and coordinate data uploads, updates and relationships with data services and publishers; probably based within Europeana Foundation	Low volume, highly relevant datasets from specialist cultural heritage publishers; bias towards museum catalogues, scholarly works and re-released classics of modern culture
Medium **(as above; additionally develop identifier resolution services)**	Larger (*e.g.* national) book data aggregators; larger retail and distribution platforms	Staffing to manage additional relationships with multiple retailers and distributors via actionable identifiers; either within Europeana Foundation or through parallel, sustainable commercial aggregator within Europeana Network	Much higher volume datasets; possibility to include wider range of scholarly and popular "trade" publications

Level of investment (cumulative)	Commercial contributor	Management model(s)	Envisaged enrichment of Europeana
High **(as above; additionally develop data and content licensing infrastructures; optional addition of federated searches to Europeana's portal)**			As above; additional possibility to link into richer metadata aggregated under more restricted licence; federated search from Europeana would enable search within aggregated "previews"

Above all, Europeana will need to take into account the needs of the commercial sector in order to create genuine partnerships. This will take one of two forms of fundamental business model. The first is that developed through the Linked Heritage specifications, crystallising around the 2010 FEP statement and comprised of 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.

The second fundamental business model envisages that commercial entities might begin to make commercial use of Europeana's data in return for contributing some of their own (though of course, since Europeana's data is available under CC0 terms, no such reciprocal arrangement is necessary). Because this type of business model goes beyond the remit of Linked Heritage, and would rest on future developments in Europeana's core service and wider ecosystem, the *D4.1 Addendum* provides further discussion, while it is only outlined here:

- Problem outline:
  - Europeana simplifies *initial discovery* of digital cultural heritage objects and a *personal research approach to navigating* among them; it does this by aggregating 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 reuse faces; 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 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 Cultural Heritage Objects and Digital Objects.
- Possible solutions:
  - Europeana could provide tools required to streamline the *commercial reuse of the Cultural Heritage Objects and Digital Objects themselves*, for example:
    - a rights discovery and 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 datasets, and syndication of rich descriptive data covered by copyright.

Several new Europeana projects suggest moves in these directions and the authors hope that this report will indicate further steps towards integration of in-copyright, in-commerce content.

1 INTRODUCTION

This report documents work building on the previous deliverables in Work Package 4 of Linked Heritage, which were:

- *D4.1 – Best Practice Report.* An introductory, technical overview of metadata best practice in the commercial sector covering:
    - Identification of products (and actors in the supply chain);
    - Description of products;
    - Data modelling (and its compatibility, through ontologies and schema mapping, with that in the heritage sector);
    - Description of some examples of existing "public-private partnerships" based on exchange of product data.
- *D4.1 Addendum.* This will respond in more detail to the changing context of Europeana's goal to redistribute its content and the EC reviewers' request to give more details of public-private partnerships aiming at commercial reuse of Europeana content or DOs discovered via Europeana.
- *D4.2 – Specification of Technologies Chosen.* A detailed implementation of best practice in schema mapping, resulting in:
    - A comprehensive mapping of the Onix for Books product information message to the LIDO metadata harvesting format, specified for mapping the latest 3.0.1 version and the previous, still widely used, version 2.1;
    - Specific technical recommendations for enhancements to the LIDO schemas, and the MINT mapping and aggregation software, to enable, respectively, the full range of ONIX's (and other commercial schemas') semantics to be preserved, and best practice in metadata management to be observed;
    - Recognition that some aspects of the metadata mapping work, particularly in regard to publication in Europeana's formats ESE and EDM, rely primarily on commercial and licensing considerations, and thus must be addressed in D4.3.

Having confirmed that aggregations of commercial data and contribution to Europeana is technically feasible, it remains to confirm the fundamental working hypothesis for a business case and explore its ramifications. This is the *quid pro quo* of delivering product description and access metadata to Europeana in return for retail links (and thus potential sales revenue for the contributor), which has been the underlying assumption since the beginning of the project. It originated in the FEP's response to the EU Green Paper A-134-O (Vaisberg, 2010), which is worth quoting in full:

> *"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."*

This approach is precisely analogous to the way that heritage organisations contribute their metadata to Europeana:

1. The Europeana portal acts as a "discovery environment" that allows the user to search metadata and previews to find *information about "*cultural heritage objects" (CHOs);
2. Part of the metadata in the Europeana portal includes a link to a richer "digital object" (DO), hosted at the provider's own website, which offers *surrogate access* to the CHO;
3. Clicking through the URL at the portal links to the DO, where the provider offers richer information, including how to *directly access* the CHO if possible.

For commercial products, the process is the same, but since there is no unique item for the public to access, the data provider simply offers *indirect access* to the product in the form of a *retail offer*:

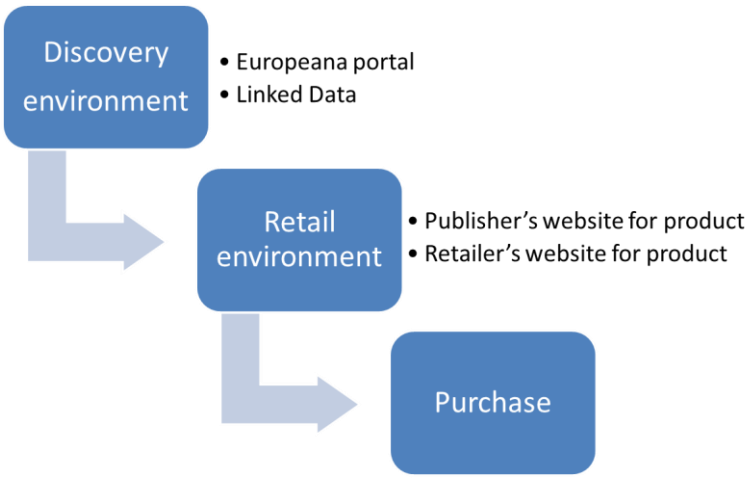


Figure 1 - user / customer experience of Linked Heritage WP4 business model

In the case of at least two of Europeana's larger contributors (Gallica[2] and [formerly] SCRAN[3]) the richer information and access provided included the option to purchase copies of books (Gallica) and license copies of images (SCRAN). Indeed, Gallica and Enclave had implemented this arrangement in national portals, as the FEP noted and as Linked Heritage D4.1 described. When Europeana introduced CC0, however, some providers, including SCRAN, chose to remove their data. Enclave notably suffered from a lack of sustained support in updating its data.

Some businesses supported directly by commercial activity run on similar models, with their own search portals, or third-party discovery environments aggregating offers from multiple retailers[4]; their funding models are dramatically different to those found in the heritage sector however, notably their reliance on advertising. These demonstrate the realism of this basic approach, and the potential for its extension to Europeana; however, serious questions of sustainability remain open.

1.1 BACKGROUND TO THIS DELIVERABLE

Even if the fundamental approach described above seems simple and clear, its practical application is not. In certain key aspects, the Europeana heritage aggregation scenario and typical e-commerce models of retail are very different:

[2] For full details see http://www.bnf.fr/en/collections_and_services/digital_libraries_gallica/a.gallica_experimentation_digital_offer.html

[3] For SCRAN shop model see http://www.scran.ac.uk/info/Buying_images.php - SCRAN no longer contributes any metadata to Europeana.eu

[4] As in the case of Find Any Film; see http://www.findanyfilm.com/about-us for details and D4.1, section 8.4.2, and also its German equivalent, MoviePilot (http://www.moviepilot.de/)

Europeana	E-commerce
One central European heritage portal for all users[5]	Multiple portals or configurations depending on 1. (perceived) audience; 2. sales rights
Metadata released under CC0 waiver of all legal rights	Metadata controlled by supply chain partners through legal (contractual or statutory) means
Display of metadata through aggregation schema (ESE, EDM)	Marketing-oriented display of data, usually using existing features of product information standard, and a set of industry norms and service level agreements
Repository–like management of metadata updates (OAI protocols)	Dynamic, message exchange architecture, allowing the metadata to track changes in price, availability and commercial strategy through a product's lifecycle
Identification at DO level (ESE) or CHO level (EDM)	Identification at product level (paid, direct access is the goal of the data, not only access to a digital surrogate or access to even richer information)
Single repository for heritage object, which also originates metadata	Multiple retailers giving access to the same product (metadata beyond a basic level may even be outsourced to a data service provider)

This report will explore in detail the problems and potential solutions which have been outlined in D4.1, and building on the practical schema mapping and software specifications in D4.2. It will follow up the research on feasible sources of data by outlining commercially acceptable contractual terms (including any further technical or organisational specifications necessary) to realise the potential in those data sources for contribution to the cultural heritage metadata corpus.

Ultimately it will assess if the commercial sector paradigm is compatible with Europeana's current DEA, and if the FEP's hypothesis based on a small number of national examples would scale to the EU level.

1.2 AIMS OF THIS DELIVERABLE

Basing itself primarily on the Linked Heritage Description of Work (DoW), this report addresses the central problems for commercial data providers and suggests solutions. Looking beyond to the wider context of Europeana's new focus on providing Linked Data for the Semantic Web, it examines some possible commercial applications for datasets published as EDM or LIDO RDF.

1.2.1 Task T4.2 – Contribution specification

The Linked Heritage description of work states:

[5] Europeana's API does enable development of new portal services tailored to specific user groups: http://pro.europeana.eu/web/guest/api - several major implementations are listed at http://pro.europeana.eu/web/guest/api-implementation - but Europeana.eu remains central.

> *"The second task will be to specify in detail how Europeana content can be contributed by the private sector. This task will focus not on technology but on the legal agreements needed to make this a reality."*

This will be addressed here under two related aspects:

1. The underlying legal framework (consisting of copyright and related legal rights) as it applies to metadata records and datasets;
2. The possibilities for agreeing licences for the access to, and re-use of, metadata.

The legal and commercial aspects of metadata are intimately linked:

- One of the primary justifications, apart from moral reasons, for legal protection of the products of intellectual endeavour, including metadata (or at least parts of it) and databases, is to ensure a return on the investment of their producers and thus encourage further endeavour;
- Commercial metadata in particular is created with two main purposes, both
  - in order to promote and market products to customers and generate sales, and
  - in order to express and track rights[6] in, and sales of, creative content for payment of royalties.

Because of their primary focus on discovery and navigation of the metadata, heritage portals like Europeana tend to place access in a secondary or even implicit relationship to the rest of the data (hence the fact that a URL to the provider's website is not absolutely mandatory, and need for revisions such as EDM with more focus on how descriptive data relate to accessible CHOs and DO surrogates).

Very loosely-defined data schemas like Dublin Core, and its ESE application profile, are adequate for full-text search and discovery[7], but have little to offer in terms of indexing power and automated, semantic navigation. The latter are not only interesting research tools, but actually essential to e-commerce, because of the need to securely identifiy – and distinguish – products and categories of products at all stages of the commercial supply chain:

> *"Analysis appears to show that there is no class, event, or extent which may not at some time be an operable value in a request, offer, or agreement. In particular, we find that modes, formats, derivations, classifications, dimensions, creation and publication dates, contributors, and their roles are all common determining factors in determining the control and flow of rights in creations.*
>
> *For example, a licencing scheme offered by an agency for JPEG downloads of any wildlife photographs by a specific photographer taken after the commencement of his exclusive agency agreement on January 1, 1996, may depend on structured descriptive metadata values for at least six of those attributes, or links, if it were to be maintained automatically, before any consideration is given to the purposes to which the picture may be put."*[8]

[6] In fact, since sales rights can restrict the specific set of customers a product can be offered to, the promotion and marketing aspect of commercial metadata clearly depends on the rights aspect: "Every [completed] transaction [or Agreement] involving creations can be described as a grant (or refusal) of rights, even where the rights are in the public domain…. The terms of the Rights Offered (and by whom) will in turn depend upon the Rights Controlled (and by whom)… the terms of the Agreements are in turn dependent on the described attributes of the Creation(s) which are being traded." (Rust, 1998).

[7] For a critical discussion of this point from another context, see http://www.rogerclarke.com/II/DublinCore.html

[8] Rust, 1998. Available at: http://www.dlib.org/dlib/july98/rust/07rust.html

Further discussion can be found in D4.1 and D4.2[9]. Briefly, the lack of these capabilities means they cannot fully express commercial data, since even where there are common information categories of interest (titles, authors, composers, performers, dates, formats…) the terms used in commercial data are so precise and (almost always) highly reconfigurable that there simply are no acceptable or useful equivalents in ESE.

Many improvements can be expected of EDM, because of its increased focus on distguishing types of entities, and linking them to events and actor roles. This is not a negative point in itself, but rather means that a significant challenge remains to be overcome for commercial data providers.

1.2.2 Other Work Package Objectives

The Description of Work also assigns several related aims for Work Group 4:

- "Rights frameworks and sharing/publishing of metadata, thumbnails, samples, *etc.*;
- Recent and/or important rights agreements;
- To outline the process whereby these technologies and facilities can be established and used to enable the contribution of private sector material into Europeana."

These formed a background to the explicit goals of T4.2 here since

a) A clear rights framework for publishing metadata (and its strongly associated marketing collateral) will be essential for commercial contributions to Europeana at scale (see 3);
b) Existing rights agreements already controlling the commercial data to be contributed must be respected; other existing agreements that can establish best practice by analogy or become useful starting points for this specific Task must be investigated and documented to follow industry expectations and avoid duplication of work (see 3.2);
c) A whole-supply chain view, including explicit reference to establishment and maintenance of technologies and facilities, will be needed to convince commercial players of the realism, quality and sustainability of the business case (see 4.6 and later sections).

Beyond the limited resources of Linked Heritage, starting to answer these needs will provide longer term value for private-public partnerships with the commercial media sectors. Contributing product metadata to Europeana under the DEA is a very narrowly defined problem, whereas sustainable partnerships need to account for rights to use or buy the full product content and previews derived from it, as seen in the national examples cited by the FEP.

1.2.3 Wider Aims of Linked Heritage

At the request of the European Commission we have also investigated potential for commercial reuse of Europeana's existing data, though this is far outside the scope of the original Description of Work either for Work Group 4 or any other WG in Linked Heritage.

While this is a very different, and certainly not directly connected, issue to the Task and other DoW aims listed above, it has been useful in presenting a well-rounded discussion of the theme of commercial involvement *per se* with Europeana and highlighting certain very serious obstacles. Please see *D4.1 – Addendum* for the full report.

1.3 METHODS EMPLOYED TO PRODUCE THIS DELIVERABLE

As in the previous deliverables, desk research (literature reviews) and expert contacts within businesses formed the basis on which to assess and present current industry best practice.

[9] Available from: http://www.linkedheritage.eu/index.php?en/142/documents-and-deliverables

Further to these fundamental activities, a small scale, relatively formal survey of attitudes was conducted, and small case studies used to gather evidence on existing rights frameworks and agreements.

1.3.1 Literature Review

As in the previous WG4 reports, the literature review is systematic but practice-oriented and informs the substantial discussion, in the next sections, of the state-of-the-art and best practice in legal and licensing frameworks.

As might be expected since we are here dealing with specifications for new services and agreements, most of the existing literature deals with the known current legal and licensing situation; for the Linked Heritage type of use case, little practical or theoretical literature exists yet, and so other methods were used, as outlined below.

1.3.2 Best Practice Case Studies

The brief case studies reported in D4.1, section 8, were supplemented with more detail from extra desk research and in some instances, informal discussion with the Work Group partners' contacts within the organisations directly involved in the public-private partnerships described. While not documented here, this provided important background for conversations with potential new commercial partners, described in section 1.3.3, and for the *Addendum to D4.1*.

1.3.3 Expert Interviews leading to Business Model Case Studies

Direct, semi-structured one-to-one interviews with experts in the commercial copyright content industries formed the starting point for moving beyond current thinking on licensing agreements for aggregation of commercial product metadata. Conversations were carried out in person (for example, at the Frankfurt Book Fair, 2012, and London Book Fair, 2013), via telephone and Skype and through email. Structure and focus for discussion was aided by making a two-fold proposal:

1. To contribute "prototype data" for ingestion into Linked Heritage and prototype publication to Europeana, given the current workflows for both projects;
2. To discuss potential ingestion of production-scale datasets through a future, enhanced Linked Heritage infrastructure and possibly future, new agreements with Europeana, documenting their requirements to make specific recommendations for future actions.

These proposals provided a concrete opportunity to define the "dimensions" of a core business model, as in 4.4, and a logical progression from a practical experience of small-scale, low-risk collaboration to specific discussions of a more ambitious, longer-term partnership.

Such production-scale contributions of commercial data are explicitly beyond the scope of Linked Heritage but could form part of the aims of a future project building on this work.

See 4.5 for the results of the Business Model Case Studies.

1.3.4 Development of Cooperation Agreement for Commercial Sector ("Prototype") Data Contributors

Through exploration of how "prototype data" can be ingested to Linked Heritage and Europeana a new contract based on the Linked Heritage Cooperation Agreement[10] was developed in order to specify two distinct but related legal contracts:

[10] For the existing Linked Heritage Cooperation Agreement for Content Providers see: http://www.linkedheritage.eu/index.php?en/182/how-to-join

1. Rights and duties of the Linked Heritage aggregation project coordinator, representing the Linked Heritage consortium as aggregator of data, on the one hand, and EDItEUR as the authorised custodian and manager, within the project, of commercial sector data contributions;
2. Rights and duties of EDItEUR, as a trusted impartial recipient of commercial prototype data, on one hand, and each separate commercial sector data contributor, on the other.

Although this may seem at first sight trivial, or even unnecessary, given that Linked Heritage already has a Cooperation Agreement accepted by many partners, and indeed, is already aggregating large quantities of data for Europeana, *this new Commercial Data Contribution Agreement is in fact the baseline for "legal agreements" to be developed with commercial partners* in direct fulfilment of the Description of Work. It is based on the cumulative collection and synthesis of all the requirements of commercial sector data contributors who engage in the interviews and follow-up discussions with EDItEUR and other WG4 partners throughout Linked Heritage WP4.

It cannot be stressed too strongly that commercial sector operators cannot be engaged in public-private partnerships without explicitly documented contractual agreements expressing *at least* these terms and conditions (see the finished agreement in section 12 and also sections 4.3 and 4.4.2).

1.3.5 Informal survey of attitudes to contribution and commercial reuse

As noted above, discussion of the central proposal of Linked Heritage WP4 allowed specific technical and legal-commercial questions to arise and these were documented as the "dimensions" of a generalised business model outline (section 4.4).

Anecdotal evidence was gathered as to the feasibility of contribution of data, and as a side issue, outside the documented scope of Linked Heritage, the potential for commercial reuse of Europeana's linked data and of the underlying DOs. These discussions are the basis of the findings of this report and also the *D4.1 Addendum* respectively.

2 LEGAL FRAMEWORK – COPYRIGHT AND DATABASE RIGHT

A legal framework underlies all commercial considerations, because firstly, it establishes the existence of categories of intellectual property and their terms of ownership, before any negotiations and transfers of that property take place.

Secondly, "copyright is granted to authors in order that they can be appropriately remunerated for their work, and so that thereby they have the appropriate incentive to provide that work …copyright itself is not an incentive mechanism, but (assuming that it is enforced) it does allow an incentive mechanism to operate, namely contracts" (SABIP, 2010). Therefore it is addressed here first as the necessary background and logical basis for considering contracts and licences for in-copyright content (potentially including metadata).

2.1 LEGAL FOUNDATIONS – WHAT CONSTITUTES CONTENT?

The types of copyright content to be discussed here can be identified in two ways: through the "lens" of Europeana's own licensing framework, and through the categories employed by commercial sector data providers. Mostly these two approaches fit closely, though with different terminology, and a few important differences in emphasis.

The Europeana licensing framework consists of four "layers", based on the "foundation" of the cultural heritage objects (CHOs) curated by heritage institutions. In the table below they are outlined together with roughly analogous equivalents in the commercial view, as in the table below:

Table 1 - commercial metadata in the Europeana licensing framework

Europeana licensing framework	Commercial analogies	Relation to full product content
Textual metadata	Textual metadata (for example, • *Books*: title, author, subject headings; • *Recorded music*: composer, author of lyrics, performers, producer, genre; • *Film and TV*: script writer, director, producer, cinematic release date, date(s) of TV broadcast; • *Photos*: caption, location, model(s), keywords)	For primarily textual products (books without illustrations) this can potentially include the full content; in practice this is unlikely. For all types of product the textual metadata will normally be partly expressed on the product itself and partly "extrinsic" (see section "Are Items of Metadata Under Copyright?")

Europeana licensing framework	Commercial analogies	Relation to full product content
Previews (thumbnail images)	Marketing collateral (for example, • *Books*: a front cover image; • *Recorded music*: album art; • *Film and TV*: a film poster; • *Photos*: thumbnail versions)	Thumbnail images typically form part of metadata for digital photographs from their creation onwards; for digital content in other media, relatively rich previews (see DO layer below) increasingly form normal content at this level.
Digital objects (DOs) – surrogates of CHOs **(in context of provider's Website)**	Richer product preview in the Web retail context (for example, • *Books*: a first chapter preview; • *Recorded music*: a whole track preview from an album, a shorter excerpt from a single track; • *Film and TV*: a film trailer; • *Photos*: larger thumbnail versions, often watermarked or otherwise technically protected)	The "retail context" here can include "substantial" extracts from the product content; for example, a book chapter; 30 second film or audio clip; lower resolution, smaller or watermarked version of the full image. What constitutes a "substantial" extract depends heavily on the nature of the product (for example, whole books, but also individual book chapters can often be separately retailed, so the sample in the latter case would be less than the whole chapter).

Europeana licensing framework	Commercial analogies	Relation to full product content
Cultural heritage objects (CHOs)	Commercial products (physical, digitised or born-digital) (for example, • *Books*: a novel or academic monograph; • *Recorded music*: an album, a single track; • *Film and TV*: a full-length movie, a TV series as a boxed set; • *Photos*: a full-resolution digital image; a printed product showing the whole image in high quality reproduction)	The product itself could be a book (print, digital or audiobook), film or TV series (part or whole) in some physical or digital format, set of one or more music recordings "released" as an album, single or other compilation, or digital photograph (in a given file format, size and resolution). Note that "the product" is often described interchangeably as though it were a single item or a class of "functionally identical" items (*i.e.* all with the same salient characteristics). All three layers based on this one refer to the product as *class of items.*[11]

In the following discussion of the fundamental legal framework, the focus is the "top" layer, the purely *textual metadata*, rather than the products and their related collateral items.

[11] See D4.2 for further discussion of the distinction between products and items: http://www.linkedheritage.org/getFile.php?id=394

Note, however, that

a) a “Digital Object” that mediates full, non-paying access to the commercial product does not exist in most cases (or is at most represented by an incomplete “preview” chapter, audio clip *etc.*), since that would remove the commercial benefits of offering the product, namely, enabling the creator, publisher, retailer and other supply chain partners to make a profit or at least recoup the costs of production, distribution and investment in more creative output;
b) the Europeana portal at present only displays thumbnail images as Previews within its search environment, but the marketing collateral potentially available in the commercial sector is far richer (in fact heritage data providers such as audio and audio-visual archives could probably offer video and sound clips too, and libraries and archives could offer full-text search, but this is currently achieved at a secondary level outside the Europeana portal itself) – and the capacity to exhibit marketing collateral and rich previews form a significant part of the attractiveness of a discovery and retail environment for potential customers;
c) there is no clear cut-off point between the full product content, richer product previews and marketing collateral – or even textual metadata where the product is primarily textual. There is no sharp dividing line because versions or extracts of, or promotions for, commercial copyright content are themselves instances of such content, though usually produced and paid for internally or in a business-to-business context;
d) here, only in-copyright, in-commerce products are considered, because of the ease with which a *quid pro quo* of cultural data for retail presence can be defined. In many cases, though, cultural bodies hold materials that are in-copyright but whose commercial status is unknown, or definitely out of commerce[12]. To integrate all types of in-copyright content, primarily the large numbers of out-of-commerce products in all media, in the heritage discovery experience, two extra factors would be required:
    i. For any given in-copyright work, an infrastructure for identifying (and ideally contacting) rightsholders in the desired in-copyright content once discovered;
    ii. An infrastructure for identifying in-commerce versions of the desired content, and, should the content be out of commerce, for communicating the request to bring it back into commerce to the rightsholders (including the required terms of access if known)[13].

The first three points, a) – c), highlight some key issues with the approaches to data aggregation, discovery environment and data republication that European has chosen. They will be explored in more detail below.

The final point d) puts Linked Heritage in the context of other EU-wide projects that cross public-private boundaries; the ARROW and ARROW PLUS projects (point d.i.) and the Linked Content Coalition (point d.ii.). These are both highly significant in that they are initiatives of the commercial sector in responding to the needs of the heritage sector while preserving the existing copyright framework, without which the commercial content industries could not survive.

[12] Formal definitions of the “out-of-commerce” state are rare, but typically, “a work is out of commerce when the whole work, in all its versions and manifestations is no longer commercially available in customary channels of commerce, regardless of the existence of tangible copies of the work in libraries [or analogous repositories for other media] and among the public (including through [*e.g.*] second hand bookshops or antiquarian bookshops)”; see: http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/20110920-mou_en.pdf

[13] Numerous references to such linkages have been raised in response to recent investigations on this topic, *e.g.* *http://www.ace-film.eu/wp-content/uploads/2012/05/ACE_statement_GreenPaper_Onlinedistribution_AV_final_111117.pdf*

The ARROW infrastructure can already be used to identify related in-commerce versions of an out-of-commerce product; these links could be exploited in future projects that aim at including commercial material in Europeana.

2.1.1 Copyright in Creative Content

The question of the legal status of commercial creative content is, for the purposes of this project, very straightforward. Only in-copyright, in-commerce products are under consideration, and they are clearly protected by copyright in all EU states[14]. Previous Europeana projects, including ATHENA[15], dealt with questions of *content* licensing extensively.

The status of full commercial products, richer previews of the product and marketing collateral falls clearly into the category of substantial copyright-protected content. Various portals and online "archives" in the Europeana project space aggregate full content and extracts[16]; analogously, commercial online retailers (too numerous to list) provide varying levels of "free access" to product content before the point of full purchase. For the purposes of this discussion such collateral materials will be considered a separate problem.

Since Europeana's adoption of the CC0 waiver of rights in all contributed metadata[17], the focus of research in the Europeana context has shifted to rights in *the metadata describing and giving access to the content.* Interest in this question historically was first raised during the growth of Open Access archives and repositories[18], which necessitated exposing large numbers of metadata records for *e.g.* academic journal articles, on the open Web, where they might be harvested *en masse* and re-used. The question has re-emerged recently due to the desire to expose metadata as Linked Open Data, as is the case for Europeana, which raises similar issues.

2.1.2 Are Items of Metadata Under Copyright?

Although it is assumed that contributors will supply large sets of metadata records, logically the first question to consider is whether each component of a record is legally protected, before turning to the collection of metadata items as a whole.

It can be helpful to distinguish types of metadata within a given record, usually corresponding in Linked Heritage's remit to a unique item (cultural heritage) or a product (content industries):

The Open Knowledge Foundation's guidelines[19] for open bibliographic data are indicative of current ideas in the open data community. They recognise two levels of data: *Core data* for "identifying" and "locating" a resource, although this appears to be basically the academic citation; and *secondary data* including all other types of description. This categorisation is illustrated by example as below:

> "*Core data*: names and identifiers of author(s) and editor(s), titles, publisher information, publication date and place, identification of parent work (e.g. a journal), page information, URIs.

[14] In accordance with the Berne Convention, 1979.
[15] See ATHENA Deliverables, especially WP6: http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents
[16] For example, EUScreen (http://www.euscreen.eu/) and EFG (http://www.europeanfilmgateway.eu/) for full audiovisual content; Europeana Libraries for full texts (http://www.europeana-libraries.eu); DISMARC for audio clips (http://www.dismarc.org/)
[17] New Europeana Data Exchange Agreement: http://version1.europeana.eu/web/europeana-project/newagreement/
[18] See Bird, 2001.
[19] See http://blog.okfn.org/2011/01/18/launch-of-the-principles-on-open-bibliographic-data/ and also http://openbiblio.net/principles/

> *Secondary data*: format of work, non-web identifiers (ISBN, LCCN, OCLC number etc.), an indication of rights associated with a work, information on sponsorship (e.g. funding), information about carrier type, extent and size information, administrative data (last modified etc.), relevant links (to Wikipedia, Google Books, Amazon etc.), table of contents, links to digitized parts of a work (tables of content, registers, bibliographies etc.), addresses and other contact details about the author(s), cover images, abstracts, reviews, summaries, subject headings, assigned keywords, classification notation, user-generated tags, exemplar data (number of holdings, call number)..."

This distinction is clear enough in terms of the *function* of the data in a minimal discovery context, but it suffers from a lack of clarity as to what the "core data" should identify; the essential creative work, or the finished publication available to end users? In terms of the heritage sector's analysis (meant for library users) in FRBRoo[20], it spreads elements of the Self-Contained Expression and Publisher Expression across the two classes of data in a seemingly arbitrary way, that does not make clear at what point in the product lifecycle each data item has been assigned, or by whom; thus losing essential clarity of ownership and provenance of the data elements. In addition, the "secondary data" category contains elements such as "non-Web identifiers" and product form ("format of work") that may be essential for disambiguation and identification of a version useful to the end user.

Adams and Davenport (2004), looking at academic journal metadata, recognise a more consistent distinction, between *a priori* metadata ("such as title, author and, generally, abstract [and also] journal, volume, page numbers, web addresses for on-line versions") and *a posteriori* metadata ("reviews"), based explicitly on the creation and publication events and their respective actors. Gadd *et al.* (2003), also analysing research publications, call their similar categories *intrinsic* and *extrinsic*:

- Intrinsic metadata:
  - Author;
  - Title;
  - Journal;
  - ISBN […*etc.*]
- Extrinsic, qualitative metadata:
  - Subject classification;
  - Authority control;
  - Indexing;
  - Abstracting.

Even though focussed on serials, this taxonomy takes more consideration of the product lifecycle, including a clear separation of the creation and publication events from the secondary bibliographic effort, and can be conceivably generalised to other types of creative content.

We make these distinctions now as they will be used throughout the rest of the report; the terms *intrinsic* and *extrinsic* metadata will be preferred as they appear to make clear the relationship of the metadata to the content (the finished product) itself.

As to whether these metadata are covered by copyright, opinions in the literature take a range of positions on a spectrum between full protection and barely any. The most prominent authors from the Open Access perspective espouse a range of views (here, in decreasing order of certainty that records are protected):

Author(s)	Opinion
Bird (2001)	"…metadata is the intellectual property of its author (even if this is not asserted in a rights statement) and the author enjoys various rights as a consequence (e.g. to be identified as the author)"

[20] Bekiari, C., Doerr, M. and Le Bœuf, P., 2010.

Author(s)	Opinion
Adams and Davenport (2004)	"If the author has not retained copyright in the paper, then it is fairly clear that the copyright in the metadata belongs with the publisher […] Where authors retain copyright it seems obvious that they grant publishers the right to use metadata along with the material"
Gadd *et al.* (2003)	"We conclude that an individual metadata record is probably protected by copyright. However the keyword here is "probably"."
	"The largest group of respondents believed that individual metadata records were facts "and there is no copyright in a fact"" [survey responses quoted in above]
Kreutzer (n.d.)	"…it turns out that copyright law usually neither protects single primary metadata nor most of the aforementioned secondary metadata. From the copyright perspective "data" in the literal sense is not protected. This applies to all mere information about the work, author, publication and so on. However complementary material can be protectable in principle. This applies especially to literary works like summaries, abstracts, reviews or graphical creations like book covers. Even classification systems can be protected (if they represent an individual/original intellectual achievement)."

Table 2- a range of opinions on the legal status of items of metadata

The first two papers cited in the table above assert that the whole set of data elements, both intrinsic and extrinsic, in a given record are always copyright protected; Bird uses the clear justification that the data's creator (or compiler, for databases) is the rightsholder, while Adams and Davenport tend to confuse authorship of the content described with rights held in the data describing it (perhaps because they also consider abstracts as being created by the original authors of the full work – often true in the academic journals sector they studied). In any case, though imprecise, this raises the question of an implied licence to use the metadata in specific ways (*e.g.* "along with the [described] material").

The latter two references in the table draw distinctions based on the intrinsic / extrinsic categorisation. Gadd *et al.* think it more likely that even the most basic facts in intrinsic data fields are protected, due to the intellectual effort of compiling them; Kreutzer, based on German law, argues that there is a sharper cut-off point between collections of "mere information" and "intellectual achievement", but this should probably be verified in the context of wider EU case law and interpretations from other jurisdictions. Finally, Kreutzer's mention of "complementary material" brings us back to the end of the spectrum where we are discussing core creative content as in section 4.1.1 above, which is undoubtedly protected. Much the same conclusions can apparently arise in USA jurisdictions[21].

[21] See for example Lubart, 2011;

It is beyond the scope of Linked Heritage to propose a solution for this question, but it must be concluded that, even accepting the least "conservative" view postulated above, the problem remains that the most interesting and valuable metadata from a cultural point of view will tend to be the more original "complementary" data which are more probably covered by copyright, leading to a tension between openness and freedom for reuse (beyond mere "discovery") on the one hand, and incentives for such reuse in the first place on the other. For Linked Heritage, and Europeana as a provider of a Linked Data aggregation, this tension is a fundamental issue.

2.1.3 Databases are Covered by Database Right

In clear contrast to the status of each individual "datum", organised sets of data (databases) enjoy definite legal status, certainly in the EU, under *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*[22].

Dierickx, B. and Vissers, R., (2009) noted in a previous ATHENA report assessing IPR in the Europeana context:

> *"In determining what rights may be applicable to a database, it is not important whether the elements in the database are copyright protected or not. The database as a whole can be copyright protected as well as protected by the specific database right."*

We cannot avoid the central point that in Linked Heritage WP4 we are dealing with deliberately created arrangements of data, usually on a very large scale, or at the very least, representing a significant portion of the output of any given publisher or other releasing organisation's metadata; these are databases with their own value as assets representing significant investment and work, as well as a certain level of creative originality based on their comprehensiveness and the quality of the data records and even individual data in context.

Curated product information has a two-fold value:

a) Firstly, especially in the Web retail environment, as marketing material, raising awareness of the product among potential buyers, and differentiating it from other similar products;
b) Supporting this, usually in conjunction with supply chain identifiers such as ISBN, it introduces efficiency and supports competition in the market for intermediary services such as distribution and wholesaling.

Hence in many developed media markets, creating and maintaining databases of product information has become a distinct business activity, albeit one that does not in itself generate exceptional profits.

In fact, the databases are not fixed repositories of records, or even growing sets of (in themselves) fixed records, but frequently updated sets of data that can be more accurately described as a core component of a data service[23]. This aspect of the data in question will form the basis of the business models discussed in section 4.

Because a large part of the value attached to a commercial metadata service comes from its currency, accuracy, and the enrichments provided by intermediary services, the question whether any "snapshot" of the data is legally protected, while remaining important, for practical purposes becomes secondary to the fact that a guaranteed data service is a clear commercial advantage –

[22] See http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML
[23] See also Linked Heritage D4.1, section 5.3.4 and D4.2, section 17.2.

according to current research, for example, demonstrably increasing sales in some cases by 100%[24].

2.2 LEGAL FRAMEWORK – CONCLUSIONS

The legal framework for contributing commercial sector product data to Linked Heritage and Europeana is straightforward, a matter of legal fact. Any agreements to this end must respect the *sui generis* database right inherent in commercial datasets, and almost certainly will also need to take into account the presence of some substantial copyright content *within* those data.

Commercial data are not primarily sourced from organisations or initiatives directly funded by the public purse and therefore the (usually valid) arguments in favour of "openly" releasing publically-funded data do not apply here, even if the subjects of the data have cultural interest and importance for the public. The data describing and giving access to those products are part of a complex supply chain that supports the creative economy, its various intermediaries and ultimately the creative work and livelihood of the original contributors themselves. Rather than being a legal matter, how the data is used comes down to questions of business value.

The next section describes how the intellectual property protected by copyright and database right in this context is currently licensed for use by Europeana and others. It contrasts the detailed agreements in the commercial sector with some of those used in the heritage sector and suggests where compromises can be negotiated in the interests of all parties.

[24] Magellan Media Consulting. (2012). *The development, use and modification of book product metadata*. Available from http://www.bisg.org/publications/product.php?p=27 – a freely available white paper is found at: http://www.isbn.nielsenbook.co.uk/uploads/3971_Nielsen_Metadata_white_paper_A4(3).pdf

3 LICENSING FRAMEWORK

To add some context to our discussion of the existing, lively market for publishers' bibliographic data, we first acknowledge that many libraries and other heritage organisations already operate in a commercial world characterised by transactions (often financial) and contracts specifying terms for reuse of their metadata.

As Bérard (2011) notes, any discussion of licensing (bibliographic or similar) product metadata must first acknowledge the already existing framework of rights discussed above, but also a substantial and complex network of contractual obligations and commercial services built on those rights. To summarise Bérard:

- There is already a commercial market for upgrading and re-selling (publicly produced) library data;
    - Hence the data (or at least, the database of all data) has a commercial value;
- Many libraries (including public bodies) charge for their metadata services (albeit often at cost price) for one or both of two reasons:
    - the on-going investment of staff time and effort to create;
    - the cost of buying (raw or enhanced) records in the first instance from commercial companies;
- For libraries opening up metadata collections, even when there is no charge, attribution is often among the terms of use[25].

Metadata used in the commercial world has a financial value – it, and the services based on and around it, such as books-in-print databases (section 5) and media listings (see D4.1), are already bought and sold, by commercial and public bodies alike. It also has a role in making work processes more efficient, and thus has an internal economic value to companies. Its role in promotion and merchandising has already been covered.

Although the World Wide Web Consortium's (W3C) report in 2011 did not explicitly mention the database right, it did acknowledge this existing licensing framework and the relevant section is worth quoting in full to demonstrate how this acknowledgement is unavoidable, even in a discussion of open data:

> *"Some library data has restricted usage based on local policies, contracts, and conditions. Data can therefore have unclear and untested rights issues that hinder their release as Open Data. Rights issues vary significantly from country to country, making it difficult to collaborate on Open Data publishing.*
>
> *Ownership of legacy catalogue records has been complicated by the degree of data sharing among libraries over the past fifty years. Records are frequently copied and the copies are modified or enhanced for use by local cataloguers. These records may be subsequently re-aggregated into the catalogues of regional, national, and international consortia. Assigning legally sound intellectual property rights between relevant agents and agencies is difficult, and the lack of certainty hinders data sharing in a community that is necessarily cautious on legal matters.*
>
> *Where library data has never been shared with another party, rights may be exclusively held by agencies who put a value on their past, present, and future investment in*

[25] See, for example OCLC's terms for reuse of its WorldCat and VIAF data services: http://www.oclc.org/worldcat/recorduse/policy/odcbynorms.htm and http://viaf.org/viaf/data/#norms (more details at: http://www.oclc.org/worldcat/recorduse/datalicensing/questions.htm)

> *creating, maintaining, and collecting metadata. Some agencies treat records as assets in their business plans and may be reluctant to publish them as Linked Open Data. Others may only be willing to release their data in a stripped- or dumbed-down form with loss of semantic detail that affects the utility of the metadata."*[26]

As the W3C "Incubator Group" noted in that study, a single dataset can contain records or even single elements merged from multiple sources, each of which may bring with it the statutory and contractual rights and existing customer and end-user terms and conditions agreed when it was integrated. Major projects such as COMET[27] described how complex these can become:

> *"Generally speaking, it's the supplier of a record who is regarded as the "owner", but the more pairs of hands (or, more pertinently, systems) through which the data has passed the trickier it can be even to determine just who the original supplier of the data actually is…*
>
> *…much of the traffic of data between systems results in new identifiers being added to records. So a single record could have a number of different identifiers, each indicative of some system's (or contributor's) "involvement" in the life-cycle of that record. It is often impossible to work out what happened when, or in what order. The record becomes, in effect, an aggregation…"*[28]

The COMET Project example above described licence chains in library data; something similar could apply in some commercial data aggregations, particularly where heritage data is sold back to commercial services for use in another context. As noted in section 3.3.3, this appears to be a growing possibility given Europeana's promotion of open data reuse.

The implication of this situation for contributions of rich data under CC0 is that guaranteeing that the rights waived by the CC0 declaration are under the control of the final contributor will be difficult, if not in some cases almost impossible. The CC0 legal code[29] underlying use of the CC0 waiver explicitly disclaims title over the contribution – so *all responsibility for checking and clearing third-party rights in CC0 data lies with the reuser* (that is, anyone providing a contribution under CC0 waives their *own* rights, but does *not* guarantee that any *other* parties' rights will not be infringed by reuse).

Unfortunately this lack of confidence in the rights position goes against the spirit of encouraging commercial reuse because of the potential cost of removing or reducing the risk of complex third-party rights claims.

3.1 IMPLIED LICENCES

As noted in Linked Heritage D4.1, section 7.2.3, commercial metadata is created for very specific use cases within a more or less well-defined community of commercial partners ("the supply chain") and comes with *at least* an implied agreement to use it only for those purposes, and not to the detriment of other partners. The example description given in D4.1, Appendix 7, taken from the ONIX for Books Best Practice Guide[30] sets out examples of allowed and restricted uses that are typical of expectations across media sectors[31], summarised below.

- Licence type: non-transferable (and usually non-exclusive)

[26] See http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Rights_issues for the exact location of this quotation.
[27] See http://www.jisc.ac.uk/whatwedo/programmes/inf11/infrastructureforresourcediscovery/comet.aspx and http://discovery.ac.uk/files/pdf/projectsPDF/discovery_project_comet.pdf
[28] http://cul-comet.blogspot.co.uk/p/ownership-of-marc-21-records.html
[29] See http://creativecommons.org/publicdomain/zero/1.0/deed.en
[30] Available at: http://www.editeur.org/93/Release-3.0-Downloads/#Best%20practice
[31] See also Linked Heritage D4.1, section 5.3.5 for more examples.

- Allowed uses relative to products described:
  - cataloguing;
  - trading in;
  - merchandising;
  - promoting;
  - selling
    - internally within the recipient organization;
    - in customer-facing applications.
- Duties relative to products described:
  - timely processing of data updates from the supplier[32];
  - compliance with (*e.g.*)
    - embargo dates;
    - valid From/Until dates;
    - announcement dates;
    - audience limitations.
- Restrictions relative to data transmitted:
  - redistribution (terms):
    - commercial;
    - non-commercial;
  - redistribution (means of access):
    - data aggregator;
    - third-party API access.
  - modification of third-party content under copyright (assuming use otherwise within the above terms).

The ONIX example above represents a maximal case covering the full range of actual practice encountered in commercial sector data; in specific cases, data will be weighted towards uses in one or another part of the supply chain (for example, rightsholder or product identification; sales reporting and royalty payments; search and display for customers in e-retail portals; identification and collation of related media assets embodying the same creative work; personal collection management data as a service) and the specific licensing arrangements will reflect this bias.

[32] Although implicit in the technical provision of update mechanisms in the ONIX specification itself, this requirement is also supported by various statements of best practice from industry bodies, such as http://www.bic.org.uk/files/pdfs/110721recipients%20best%20practice%20final.pdf and http://www.bisg.org/publications/product.php?p=27

3.2 SPECIFIC END-USER LICENCES

The understanding of implied licence to use ONIX data in section 3.1 is not an isolated example produced by one data provider; rather it represents a synthesis of terms and conditions commonly found in the commercial sector, aimed at providing orientation for ONIX users new to the market.

To illustrate this fact, the table below presents representative extracts – only *illustrative* of the full licences, found in the footnote links on these pages – for representative data services in each sector. They are not intended to be fully comprehensive, since each specific service will meet different user needs and emphasise different types of data content.

Some of the examples presented here are services provided by non-commercial, or at least primarily non-profit companies; more commercial examples are found among the case studies in section 4.5. Note that *all* impose *some* restrictions, and that most are *more* restrictive than the implied licence above.

Table 3 - real-life metadata reuse terms of books, AV, music and photo services

Sector	Service	Licence type	Allowed uses	Duties	Restrictions
Books	LibraryThing APIs[33]	Non-exclusive (implied by the fact that they are APIs)	Various; all have some combinations of the duties and restrictions listed here	Various, including, for some options: Attribution (notice, [hard] link or both);	Various, including, for some options: No commercial use; Limited API hits (especially where an API key is needed[34])

[33] Full terms at: http://www.librarything.com/wiki/index.php/API
[34] See API key licence details at: http://www.librarything.com/services/keys.php

Sector	Service	Licence type	Allowed uses	Duties	Restrictions
Film & TV	IMDB[35]	"Limited, non-exclusive, non-transferable, non-sublicensable"	"Use, reproduce, publish, distribute, and display any IMDb Title ID or an IMDb Name ID"	Payment of fees Update "at least once every 10 days" display the IMDb logo link to IMDB site	Take down within 24 hours if "risk" detected by IMDB
Music	Universal Music UK Artists Gateway API[36]	"Limited, personal, non-exclusive, non-commercial, revocable, non-assignable and non-transferable"	"Display the AG Content on your website and/or software application"	Warrant not to use on websites or apps that defame, harass, deceive, threaten *etc.*; Keep intact all links;	No commercial use ["without obtaining a written licence"]; No sub-licensing; No modification;

[35] Full terms at: http://www.imdb.com/licensing/subservicetc
[36] Full terms at: http://umusic.co.uk/artists/use-of-our-data

Sector	Service	Licence type	Allowed uses	Duties	Restrictions
Photos	Flickr[37]	"Non-exclusive, non-sublicenseable"	[Not specified]	Comply with rights statements of each image (as well as updates to these); Remove content upon request by users; Use specified text for (prominent) attribution; Comply with terms and conditions specified by Flickr within *your* site or application	No commercial use without a special API key; No service replication ["any application that replicates or attempts to replace the essential user experience of Flickr.com"]; API hit and result display limits; No sale or sublicensing of API

[37] Full terms at: http://www.flickr.com/services/api/tos/

3.3 CREATIVE COMMONS LICENCES

Described here because they are widely known and used in the heritage sector, notably by Europeana, and hence forming the context for Linked Heritage's work on licensing, these are a set of 6 modular licences[38] designed to encourage re-use of creative cultural, technical and scientific content within the limits provided by existing copyright, by making the terms and conditions for (re)use clear and comprehensible.

Europeana's "Data Exchange Agreement" uses a Creative Commons formulation to go beyond a conventional licence to waive all rights in the content (with a "Public Licence Fallback" to cover jurisdictions where this is not legally possible), applying this specifically to the textual metadata it releases as Linked Open Data[39].

3.3.1 CC Licences: Reproduction, Distribution, Attribution, Derivation, Commercial Use, Like Terms, and Licence Notices

The possible terms of use available in the CC licences are outlined below:

Abbreviation	Duty	Restriction
BY	Attribution: Credit be given to copyright holder and/or author	
ND		No derivative works
NC		No commercial reuse
SA		"Share alike" [all derivatives or copies must be distributed under the same terms as the original]
Notice	Copyright and license notices be kept intact	
Reproduction		Making multiple copies is allowed

Some of these terms coincide with those in the commercial sector, but notably they are far more general and open to wider interpretation. They are also far less service-specific; and none of them establish a direct, explicit ongoing relationship between the user and creator (or supplier) of the data. They are therefore of questionable value for establishing a commercial partnership in the usual sense, since they obviate the need to (or the possibility of) injecting an agreed, legally guaranteed value (financial or in kind) into the supply chain in return for goods or services rendered.

3.3.2 CC0 1.0 Universal (CC0 1.0) - Public Domain Dedication

The Creative Commons "tool" for explicitly dedicating known copyright works to the public domain is called "CC0" and is properly a waiver of rights, including a declaration that no legal recourse for

[38] Details of the licences are at: http://creativecommons.org/licenses/
[39] See also http://wiki.creativecommons.org/CC0_use_for_data

infringement of any inalienable rights will be made, plus a "fallback" option offering a universal public licence in the case that the waiver has no legal force[40]. Relevant points for Linked Heritage and especially commercial data providers to note are:

- The CC0 instrument specifically waives *both* copyright *and* database rights:

  "CC0 is intended to cover all copyright and database laws, so that however database rights are protected (under copyright or otherwise), those rights are all surrendered."[41];

- Although Creative Commons draws a distinction between "marking" a work already known to be public domain and dedication through a CC0[42], in practice there is a problematic condition for *users of the dedicated content*:

  "CC0 contains a disclaimer… so there is no assurance whatsoever that [the person who applied CC0 to the work] has all the necessary rights to grant permission to the CC0'd work… or has cleared any uses of third-party content. [Creative Commons] recommend that you do not use the work until you have taken all... steps and precautions… which may include contacting the person who applied CC0 and consulting legal counsel."[43]

Here is the full text of the disclaimer in the CC0 Legal Code[44]:

> *"Affirmer disclaims responsibility for clearing rights of other persons that may apply to the Work or any use thereof, including without limitation any person's Copyright and Related Rights in the Work. Further, Affirmer disclaims responsibility for obtaining any necessary consents, permissions or other rights required for any use of the Work."*

This disclaimer pushes back the question of the ultimate ownership of data to the final reuser of the data. As noted already in the introduction to 3, commercial data and even some heritage organisations' data already contains subsets of records licensed from other providers and these third-party rights should be cleared. *To reuse CC0-marked data in compliance with database right and existing licenses, thus mitigating the risk of future rights claims, implies difficulty and probably associated costs.*

Therefore taken as a whole without any additional guarantee of up-front rights clearance, *the CC0 waiver has little or no value at all for enabling reuse of metadata, and may in fact make it more difficult*. It may be misleading for the data reuser, in that by applying the CC0, Europeana may *appear* to have made these checks, or required them of its contributors, when it is under no obligation to do so.

Further, in focussing primarily on the availability and reuse potential of metadata, Europeana avoids an important issue, that of licensing and reuse of the Digital Object and Cultural Heritage Object (or commercial products and their marketing collateral) which remains the fundamental source of commercial value in existing supply chains. However, other projects involving collaboration between the heritage and commercial sectors do address this issue (see 2.1) and are starting points for a genuine way forward to reuse of copyright content.

[40] For the full wording of the CC0 waiver, declaration and licence, see http://creativecommons.org/publicdomain/zero/1.0/deed.en
[41] *"What are database rights?"* from http://wiki.creativecommons.org/CC0_FAQ#What_are_database_rights.3F
[42] See http://wiki.creativecommons.org/CC0_PDM_comparison_chart for a detailed comparison according to Creative Commons
[43] From http://wiki.creativecommons.org/CC0_FAQ
[44] http://creativecommons.org/publicdomain/zero/1.0/legalcode

3.3.3 Potential issues with fully open access metadata sets

Finally, since the earliest days[45] of open publication of large description and access metadata sets it has been acknowledged that intellectual property issues could arise that would challenge the business models of even non-commercial or non-profit data providers. The following "issues", identified[46] in the context of Open Archives (OA), could also apply to the entire Europeana aggregation, or specific, identifiable datasets within it, in the absence of clearly defined and enforceable restrictions and obligations for reuse:

Issue	Example
Forks with lockout:	A takes B's metadata, cleanses or enriches the data, sells a service, and refuses to let B (or B's community) have the improved data.
Stealing credit:	A takes B's metadata, puts their own name on it, and registers a new data provider with the same content (or a new service provider), not crediting B.
Brand name:	A takes B's metadata, messes with it [that is, modifies it and introduces errors], and provides a different service, using (and tarnishing) B's respected name.
Claw-back:	A takes B's metadata, sells a popular commercial service, and B subsequently modifies the rights statement to exclude commercial use, and threatens to sue A if A doesn't remove B's data from the service.[47]
Liability:	A takes B's metadata, which B had collected from many sources including C. A sets up a commercial service. C sues A for violating its IPR, and A holds B liable since B didn't transmit C's IPR statement correctly.

Table 4 – Quoted in full from Bird, 2001, "A Survey of Intellectual Property Issues Relevant to Metadata"

The author of the above "survey" also notes that "This list is very incomplete." See case study 4.5.10 for a real-life commercial example that could easily be added.

As a basis for a business model of open archives, the CC0 structure at first sight seems simple. In detail it may well be far from that:

- in the first four issues above, because of its contrast with Europeana's stated usage guidelines[48] and varying applicability in (and across) national jurisdictions[49], and;
- secondly (as in the final "liability" issue above) because of the existence of composite metadata accumulated from many sources and editing iterations.

[45] See the outline history of the Open Archives Initiative at: http://www.openarchives.org/documents/jcdl2001-oai.pdf
[46] See Bird, 2001.
[47] This issue would only be relevant for updates to the dataset after the initial publication under CC0, since the original public domain dedication and universal licence are irrevocable – but note that the value of commercial data-as-service lies in its currency and comprehensiveness (as discussed in section 4.4.4).
[48] See section 3.3.7 and recommendations and community norms at http://pro.europeana.eu/web/guest/licensing
[49] For the extra complexities for data creators, aggregators and reusers, see http://wiki.creativecommons.org/CC0_FAQ#Why_do_some_works_indicate_the_jurisdiction_from_which_the_work_is_being_published.3F

If a mistake is made at the aggregation stage it could lead to significant complications for the reuser, where, in the case of commercial metadata, binding legal agreements and sustainable commercial services already exist above and beyond the limiting case of an OA archive. This will be discussed in more detail in section 4.

Europeana's DEA commits Europeana to remove data from the portal on request from the contributor. This does not remove the risk of any of the above "issues", because any rights over data that may have been extracted by third parties in the meantime have already been waived, and on the other hand, CC0 is explicitly irrevocable:

> *"[the CC0] Waiver shall not be subject to revocation, rescission, cancellation, termination, or any other legal or equitable action"*[50]

3.3.4 Linked Heritage Response to the Europeana DEA

In order to give more flexibility and choice to its partners, the Linked Heritage consortium formed a task force to respond to the Europeana DEA upon its announcement in 2011. A filtering option at the data publication stage was added to the MINT aggregation server so that three levels of record completeness can be selected:

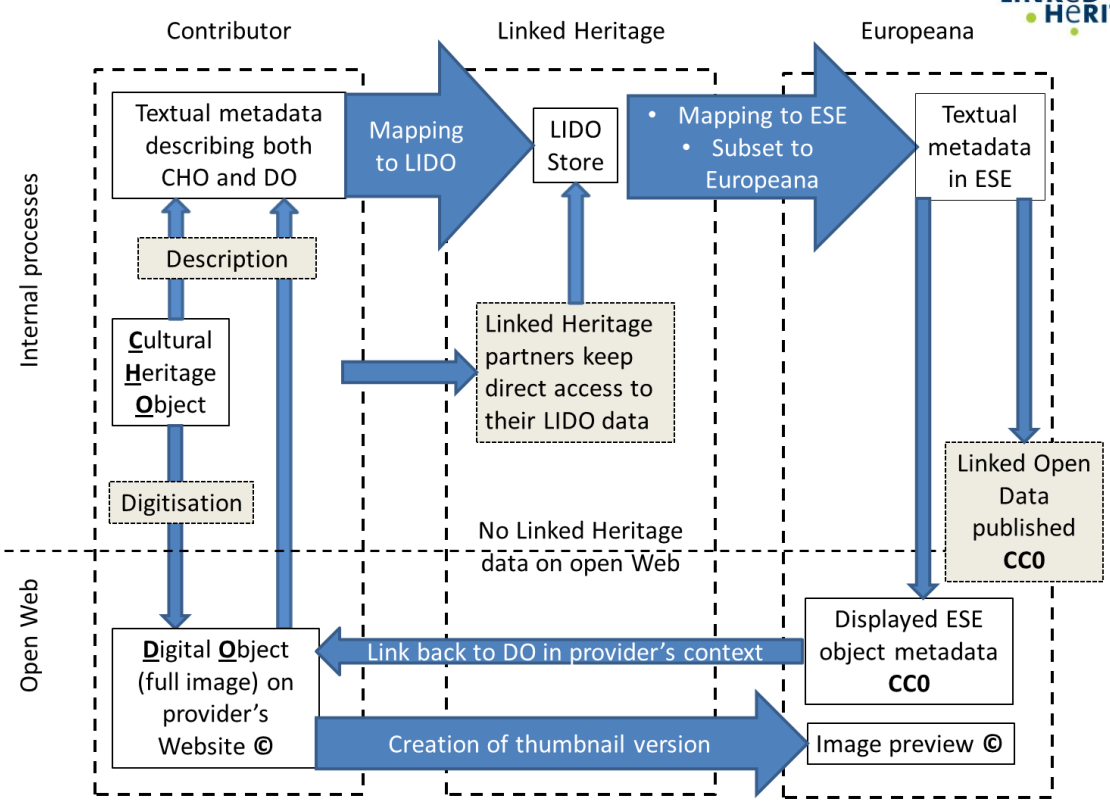
1. Mandatory only: includes only the mandatory elements both of LIDO and of ESE in the data supplied to Europeana.
2. Intermediate: includes all available data elements except those mapped to ESE's dc:description elements; ensures that the most valuable (possibly copyright protected) textual content is omitted from the data supplied to Europeana.
3. Full: includes all available data elements that can be mapped to an ESE element in the data supplied to Europeana.

These categories can also apply to commercial data contributions. WG4 defined in the same document the level of contribution expected from commercial sector contributors in terms of the specific use to be made of data supplied to Linked Heritage. The distinction is between

1. "Test" data used *only* internally to Linked Heritage, stored in the LIDO format;
2. "Prototype" data – subsets of the "test" data, translated to ESE format and published through Europeana.

Figure 2 shows these two schema mappings in the context of the whole Linked Heritage / Europeana workflow. Further discussion of this distinction can be found in Linked Heritage D4.2, and the concepts were defined further in creating the contractual proposal used in the case studies for D4.3, where they appear as Data Contribution ("test") and Publication Subset ("prototype").

[50] http://creativecommons.org/publicdomain/zero/1.0/legalcode



3.3.5 Current Linked Heritage Aggregator Configuration

Before discussing Europeana's licensing framework it is useful to review Figure 2 above to clarify the terms used for the types of content involved. The diagram shows the complete aggregation workflow from contributor, through Linked Heritage, to Europeana, distinguishing between what is available on the open Web and what is internal to each partner in the process.

Note that, due to the Linked Heritage filtering option described in 3.3.4, although rich data is stored in the LIDO server within Linked Heritage's infrastructure, not all of this must be

Figure 2 - Linked Heritage and Europeana aggregation workflow

contributed to Europeana. This distinction is carried through to the Commercial Data Contribution Agreement (see 12) developed by WP4, with the distinct concepts of Data Contribution ('Test') and Publication Subset ('Prototype') (see 4.3).

3.3.6 Types of content and rights covered by Europeana's DEA

To summarise the picture of Europeana's licensing framework we now revisit the practical definition of "metadata" in Europeana's context. The table below illustrates the distinctions currently drawn by Europeana, adding one detail from the cultural heritage perspective that was implicit in the original diagram (Keller, 2011)[51], namely the detailed description ("documentation"), normally created by the custodian institution of the cultural object. This is "[t]he documentation of collections [which] includes the detailed description of individual items within collections, groups of items and collections as a whole… contextual information: the historical, geographical and theoretical background that gives museum collections much of their cultural significance and value"[52].

"Description" in this context will include "free text" such as would typically be found in metadata fields such as dc:description (and its subelements) in ESE/EDM; the "free text" content may well be copyright protected. However, as is clear from the quote above, it also includes detailed, structured, typed data, collections of which may attract database right.

The most problematic layer is clearly the top "metadata" layer, where normally specific end-user licenses exist, illustrated in 3.2. However, with a shift in emphasis towards more granular and machine-readable description of rights throughout the framework, this incompatibility could be overcome. Otherwise, it seems likely that only the most minimal factual data will be included in any "Publication Subset".

[51] See http://www.slideshare.net/paulkeller/main-aspects-of-the-europeana-data-exchange-agreement
[52] See introduction to the CIDOC-CRM definition, found at www.cidoc-crm.org/

Viewpoints	Layers	Content at this layer	Content use by Europeana	Rights in this layer's content
Europeana	Metadata	Textual data describing the DO and CHO, in ESE format, presented in Europeana's search results. Includes **links** to DO and preview.	Available for search and discovery in Web portal to any browser, anywhere. Published as whole dataset "dumps" or "snapshot" downloads (EDM in RDM format). Available via API.	• Europeana's metadata guidelines (not binding) • CC0: - Waiver of all rights - Universal licence for any kind of use - Promise not to exercise any remaining legal protection
	Preview	The low-resolution version of the DO for use in Europeana's Web portal.	Europeana displays the preview in its portal, and publishes the link to the image on its server under CC0.	The same rights as the DO; rights statement appears on portal but not directly in the image file.
Both	Digital Object (DO)	The digitised version of the CHO, available through the cultural heritage institution's Website.	Europeana only displays a Web link to the DO in its portal, not the full DO.	Rights assigned to the digital image, depending on national and EU copyright law, and licences decided by the cultural heritage institution.
Cultural heritage institution	Provided Cultural Heritage Object (CHO)	The physical object in the collection of a museum, gallery, archive or library.	No direct use of the actual object by Europeana.	The rights of the legal custodian(s) of the object are unaffected; they may be the basis for rights in the DO in the layer above.
	Description / documentation	The detailed description and explanation of the significant of the CHO	This may be (partly) included in the metadata provided to Europeana	Depending on national copyright laws, this may be a substantial piece of creative work protected by copyright.

3.3.7 Current Europeana Licensing Framework

Europeana's new Data Exchange Agreement (DEA)[53] was officially adopted 12 September 2011. The transition schedule announced required the signature of all direct contributors of data to Europeana by 31 December 2011, and data supplied by contributors which have not signed has been removed from Europeana. It is required of data aggregators as well as organisations and individuals who directly contribute their own data. This therefore also applies to Linked Heritage, and for the purposes of Work Package 4, any and all commercial organisations who might contribute data.

The DEA amounts to agreeing to allow Europeana to publish contributions of metadata under the Creative Commons "CC0 waiver"[54], offering the maximum legally available freedoms to any party who wants to reuse the published data. The terms applicable to reusers of data published by Europeana under the DEA and the accompanying, non-binding "usage guidelines for Metadata" (draft, 2011)[55] are compared below:

Table 5 - detailed analysis of Europeana's non-binding reuse guidelines

Possible restrictions and obligations	Data Exchange Agreement	Usage Guidelines
Reusers attribute data creators and providers	No obligation.	*Should* attribute creators and providers, linking to them if possible. *Should* include rights information directly with data, or at least linked to it.
Reusers' updates to metadata	No obligation.	*Should* use API, or at least link back to Europeana to allow discovery of updates there.
Reusers' modification of data	No obligation or restriction.	*Should* make clear where reusers have changed data since publication in Europeana. *Should* republish modified data under the same terms as Europeana published it.
Reusers' representation of provenance of data	No obligation or restriction.	*Should not* represent republished data as authoritative or used with endorsement or permission.
Reusers' relation to EU and/or national laws	CC0 defaults to unlimited universal licence and/or waiver of any legal rights or recourse in relevant jurisdiction.	*Should not* breach EU or national copyright, defamation or data protection laws.

[53] All details of the new Europeana DEA cited in this section are taken from the official DEA page at http://version1.europeana.eu/web/europeana-project/newagreement/
[54] For Creative Commons' own explanation of CC0 see http://creativecommons.org/about/cc0
[55] See http://pro.europeana.eu/usage-guidelines

Even though the guidelines make *recommendations* for the responsible reuse of Europeana's data, Creative Commons' own description of CC0 makes clear that *no legal protection is possible* once the data is declared to be in the public domain "to the fullest extent permitted by law", and "anybody will be able to use your work for any purpose, even in ways you may find distasteful or objectionable. They can also make money off of your work, and they may give you credit or they may not"[56].

Finally, the Europeana DEA and Guidelines could possibly lead contributors to believe that they retain some control over copies of their data through correction or removing it from the Europeana portal and data services (Europeana DEA Articles 2.5. and 7.3.), whereas Europeana has under the same DEA (Article 3.2.) committed to release its data *in perpetuity and irrevocably* under CC0, and therefore can give no guarantee that anyone, including *Europeana itself* may not simply retrieve any subsequently withdrawn data thus published (*e.g.* from a third party who harvested the data under CC0 while available from Europeana) as that data *is still legally and effectively in the public domain.* The implication is that the only effect of removing data from Europeana would be to prevent updates and corrections of the data from being distributed under the same terms as the original material.

3.4 LICENSING FRAMEWORK - CONCLUSIONS

It should be clear from the considerations in this section that the existing licensing arrangements for commercial data sets are not immediately reconcilable with Europeana's CC0 condition. Of course, were Europeana to introduce other rights frameworks for commercial data, there will be a far wider choice of licensing models for discussion. However, this report assumes no change to the DEA.

Between the extreme options of contributing no data at all to Europeana, and releasing a company's entire data assets under the CC0 waiver, there may be a spectrum of options involving further specification of the proposed exchange when the costs and benefits to each party are better understood – this corresponds to a practical implementation of the findings of the value chain and content development strategy analyses presented here.

Linked Heritage is in an excellent position to work out these nuances, having already created a detailed, flexible technical solution for our heritage sector partners, and also having partners directly involved in the commercial data supply industry, at various stages of the supply chain:

- EDItEUR – international standards and best practice;
- NSL, mEDRA and MVB – identifier registries;
- MVB - books in print data services.

3.5 LICENSING FRAMEWORKS – METADATA OR CONTENT?

One final point can be made in the context of licensing frameworks, looking beyond the scope (and available resources) of the Linked Heritage project. As has been stated more or less explicitly since the beginning of the Linked Heritage project, metadata for commercial (and also many non-commercial, heritage sector) products does not exist independently from the need to respect statutory rights *in the products themselves and their content*, nor is it irrelevant to solutions for obtaining and using *contractual rights* in the products and their content. This can be summarised by three essential points[57]:

[56] "Can I control how my work is being used once I publish it using CC0?" from http://wiki.creativecommons.org/CC0_FAQ#Can_I_control_how_my_work_is_being_used_once_I_publish_it_using_CC0.3F

[57] All based on statements from Rust, 1998; available at http://www.dlib.org/dlib/july98/rust/07rust.html

1. "Every transaction involving creations can be described as a grant (or refusal) of rights, even where the rights are in the public domain"; because "in operational if not commercial terms, there may be little or no difference between the digital transactions of an e-publisher and an e-library."
2. "…in e-commerce, rights metadata is dependent on descriptive metadata" because knowing what rights you have (or want to request) depends on identifying *what those rights apply to*, through description, classification and relations;
3. "the reverse is true" and there is an essential "dependence of Descriptive metadata on Rights metadata" because "the rights transaction is able to generate automatically a new descriptive metadata set through the interaction of the agreement terms with the original creation metadata". This will be especially true in the case of:
    a. requests for, creation of and acquisition of *new versions of existing content* as envisaged in 2.1;
    b. reuse of *datasets as content* as in 1.2.3.

So, not only is *addressing full content access* (*i.e.* licensing full object content) ultimately inevitable when looking at metadata exchange, but also *descriptive richness and reuse flexibility.* Good "rights data" has the same qualities as good "descriptive data" once a certain critical mass of content or large enough scale of reuse is reached.

Europeana is currently starting to address these points through

1. Upgrading their ESE/EDM schemas[58] from reliance on Dublin Core[59];
2. Dissemination of best practice[60] in rights labelling[61];
3. Investigation of alternative, granular rights expressions for data and DOs[62].

We should aim to progress beyond this point towards a truly e-commerce enabled ecosystem, because "E-commerce offers the opportunity to integrate the functions of discovery, access, licensing and accounting into single point-and-click actions in which metadata is a critical agent, a glue which holds the pieces together" (Rust, 1998); experience shows that existing technology itself can provide solutions to problems it may have been used to create[63].

From the current focus on *datasets as content* and the apparent usefulness of CC0 in short-circuiting the need for detailed rights expressions[64] to enable content reuse, we should shift

[58] *E.g.* through the EDM-FRBRoo task force; see: http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+EDM+FRBRoo
[59] "Dublin Core's structure is neither tight enough to satisfy the requirements of a rights-based system (which as we have seen needs a fully structured framework) nor loose enough to be able to accommodate such a structure being imposed from another source… At each stage in its progress, Dublin Core leaves open more doors for exceptions and anomalies… its attributes, terminology, and examples are derived principally from text-based creations… despite the fact that a cross-sector vocabulary is now essential. [And] it views rights metadata as an extra (15th) element or set of elements, not recognizing that, in fact, it embraces 13 of the other 14 elements (Titles being the only exception)." Rust, 1998.
[60] For example, about the inclusion of rights information in core metadata, hence the figure of less than 50% of DOs carrying a rights statement, and lack of persistent identifiers for CHOs and their DOs; see http://pro.europeana.eu/documents/866067/983522/MS11+Collections+and+Data+Analysis%2C+Strategy+and+Plan+2012-2014
[61] See http://pro.europeana.eu/web/guest/pro-blog/-/blogs/1442749
[62] Through the eCloud project: http://pro.europeana.eu/web/europeana-ecloud
[63] "The Answer to the Machine is in the Machine"; see http://media.wix.com/ugd//bff7bc_1faf4c950f2df2b17592b895f46bab6d.pdf
[64] See, for example, http://rhizomik.net/html/~roberto/thesis/html/RightsExpressionLanguages.html

attention towards *the core needs in relation to the Digital Objects (and the underlying Cultural Heritage Objects and commercial products)*, which are:

1. simple access to licensing options, contact with rightsholders *etc.*;
2. automated clearance options, including payments.

In this way we could see reuse of primary content (including datasets, but mainly the rich content (the DOs) they refer to) dealt with reliably, securely and at scale, with a multitude of attractive commercial applications.

The *D4.1 Addendum* deals with these issues in more detail, and relates them to Europeana's new projects aiming to enable licensing of content and data[65].

[65] Especially Europeana Cloud: http://pro.europeana.eu/web/europeana-ecloud

4 BUSINESS MODELS AND SUSTAINABILITY

The business model proposed by Linked Heritage WG4 has already been presented in the introduction (section 1). Here its ramifications will be worked out in detail, with reference to normative documents from Europeana and Linked Heritage, and practical considerations *based directly on dialogue with prospective commercial partners*.

The existing Linked Heritage and Europeana frameworks have only been used as a starting point, with the expectation that, given the tensions identified in the preceding two sections (see 2.2 and 3.4), indications will be given at relevant points in the discussion as to how to go beyond the current framework.

4.1 QUID QUO PRO – RETAIL LINKS FOR PRODUCT METADATA

The Linked Heritage public-private partnership is based on the assumption that supplying product data for culturally interesting and valuable commercial creative works is worthwhile for commercial publishers, retailers and data aggregators because they derive the benefit of greater exposure and potential incremental sales of those products (through retail links provided in the data) to balance the costs of providing the data.

4.1.1 Cost and benefits beyond the basic exchange model?

There may be other (potential) costs and benefits associated with this exchange:

	Europeana	Commercial partner
Costs	Overheads of maintaining data repository: - Relationship management; - Data format transformation support (or adoption of up-to-date industry standards); - Updating data feeds and correcting errors	Overheads of maintaining data feed: - Relationship management; - One-off cost of adopting industry standard; - On-going costs of updating data formats to latest version of standard - On-going provision of updated data
Benefits	Increased data richness: - Better coverage of cultural topics; - More links between objects; - Increased linked data reliability (based on robust commercial data models); Better understanding and cooperation with commercial sector through collaboration Association with high-profile commercial endeavours	Increased sales (potential), especially for - High retail priced items; - Backlist items; Association with high-profile European cultural endeavours

In the process of raising awareness of this proposal among commercial sector companies and specifying the technical and commercial requirements to make it a reality, other variations on this model, together with other potential costs and benefits, have been discovered. They are

documented below at the points where alternative opportunities became apparent, especially in 5 and 6.

To illustrate the more specific needs of the heritage and commercial "supply chains", the example of e-book *content* and *metadata* supply chains will be described in section 4.1.2, since here **a thriving commercial market for product metadata already exists**, and served by a variety of companies, whose business models are commercial to a greater or lesser degree, depending on the relationship of metadata creation or services to their core business.

4.1.2 The content and metadata value chains in European e-book markets[66]

We have already compared metadata supply chains in heritage versus commercial contexts[67]. However, in order to illustrate costs and benefits to specific partners in these supply chains, we must compare them with the respective content supply chains. Partly this is because, as previously mentioned, there is no hard and fast distinction between "metadata" and "content", neither in heritage or commercial contexts; but primarily because **the value of metadata to any partner depends – directly or indirectly – on some value (be it financial or otherwise) derived from the content the metadata identifies and describes, and to which it provides access.**

For many good reasons, the Europeana model was designed to serve the European cultural institutions, increasing visibility and discoverability of their collections. This implied a simple value chain: users access the full content directly from the cultural institution site, without intermediaries.

Europeana operates in this short value chain aggregating metadata[68]. So we have two value chains: one, extremely short, for content; the other, more complex, for metadata (see Figure 3):

Figure 3 - Content and metadata value chains in the Europeana model

CI = Cultural Institution EU-Ag = Europeana aggregators

From the user's perspective (right to left), the supply chains work as follows: Users search for content in Europeana and retrieve what has been described through appropriate metadata by cultural institutions (CI). To make this happen, cultural institutions provide metadata to

[66] This section (4.1.2 as well as 5) contains substantial contributions from mEDRA, a Linked Heritage partner with considerable experience in these areas.
[67] For example, in D4.1, section 5 (available at http://www.linkedheritage.org/getFile.php?id=283) and in sections 2.1 and 4.1.2 of this report.
[68] Technically, Europeana operates very often through intermediaries, of which Linked Heritage and ATHENA consortia are some of the largest; the progressive integration of such intermediary aggregators within the Europeana Network indicates that they can be considered an extension of Europeana's activity for this analysis (but see section 8 and the *D4.1 Addendum* for some exceptions).

Europeana either directly, or – more commonly – through aggregators, which act as intermediaries in this chain.

However, *this does not lengthen the content value chain*. Once users find the content of their interest, they access it directly at the CI website, using a URI that has been provided to and displayed in the Europeana system. *Only the metadata value chain is longer*, and more complex, since Europeana is an additional data source for the user, which does not substitute for the possibility that the user might search the CI's website directly[69].

Note that *there is no reverse data flow* – this would be the (not yet attempted) rights chain alluded to in 2.1.

When we move from cultural institutions to commercial content providers, Europeana starts interacting with a more complex value chain, which includes several intermediaries between the producers and the users. If we look in particular at the e-book value chain, we have a picture like that described in Figure 4:

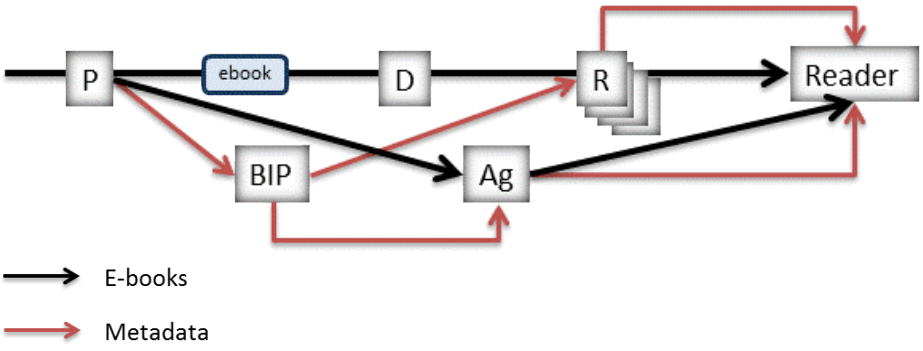


Figure 4 - Content and metadata value chains in the e-book market

P = Publisher

BIP = Books In Print database

D = Distributor

R = Retailer

Ag = Aggregator

Looking at Figure 4, right to left in terms of value, the process starts when a reader searches for and discovers an e-book in the website of one Retailer (R), like an Internet bookshop, or of an aggregator (Ag), as it is frequent in scholarly publishing. This is made possible by the fact that retailers usually collect rich metadata about e-books (including, for example, an abstract, a thumbnail of the cover, a description, a table of content, *etc.*) from bibliographic agencies that produce Books In Print (BIP) databases. Bibliographic agencies, in their turn, collect data from Publishers (P), and provide various services including data quality and enrichment, format standardisation, *etc.*

Once the required e-book has been selected, it is purchased from the chosen Retailer and downloaded in one or more devices (including PCs, e-readers, smartphones, tablets). To allow digital purchase, Retailers have a contract with the Publisher that authorises them to sell the e-

[69] Financial compensation for the creation of the digital content in these value chains comes in almost all cases directly from the national State, or, in some cases through the EC or other supra-national funding bodies, and in a relatively small number of cases, through a token payment by the final content users. Public-private partnerships can also play a role in the creation of Digital Objects (see 4.4.1) but interestingly for Linked Heritage, metadata are rarely created separately from the DO, and certainly not by others than the original CHO's curators. This contrasts strongly with commercial data creation, which is commonly outsourced.

book. To deliver the file, Retailers get the file directly from the Publisher or from a Distributor (D). Again, the reverse chain of orders and sales information is not shown, for clarity.

Such digital distributors replicate to a certain extent the similar role in the print books supply chain. This is the case in most of the continental Europe. For example:

- In the Netherlands, CB Logistics, an organisation also dealing with distribution of physical books, and which also produces the Dutch books-in-print database, entered the e-book distribution market[70];
- In Italy, some new companies have been created to act as digital distributors. The largest initiative is Edigita (www.edigita.it), jointly created by some of the leading publishing groups; but also other newly created companies entered the market providing services in particular to small publishers and acting also as retailers serving end users, something that Edigita doesn't cover;
- A very similar model is growing in France with a platform like Eden Livres (www.edenlivres.fr), created by a group of medium-sized publishers but – like Edigita in Italy – open to other publishers;
- In Germany, www.Libreka.de, an initiative launched by MVB also covers different roles in the chain: as distributor, retailer, and aggregator. Moreover, MVB itself produces the books in print catalogue for the German speaking area (VLB);
- Based in Ireland, ePub Direct (www.epubdirect.com) provide a distribution service to publishers at global level.

This brief, non-exhaustive, description of the evolving landscape of e-book distribution in Europe provides a glance of the complexity of the value chains involved. In the synthetic description of Figure 4, *one commercial operator frequently plays more than one role, depending on the historical development and current status of the markets and other companies involved*.

Planning the integration of commercial items in Europeana needs to take into account this complexity. Three questions arise:

1. Who could be the best provider for high quality metadata for commercial e-books into Europeana?
2. To which resource(s) should a metadata record registered in Europeana resolve, so to allow users to get the full text and possibly additional information about the content?
3. Which impact the open metadata policy of the DEA has to this picture in terms of incentives to possible private partners to join Europeana?

The first two of these questions are approached in the summary findings (sections 5 and 7) of this report, and the concluding section, 8. The problem of the Europeana DEA is addressed separately, in section 3.3 and in the *D4.1 Addendum*, where it is placed in a broader context of open data business models.

Normally for Europeana, a data contributor (an organisation) will be the curator of the CHO, creator and owner of the DO, and maintainer of the "full information context", including its textual

[70] CB Logistics (www.cb-logistics.nl), formerly Centraal Boekhuis, is a company owned by the publishers and booksellers associations in the country, which also acts as bibliographic agency, so covering more than one role in the described value chain. The strategies adopted when entering the e-book distribution have been described by Hans Willem Cortenraad, "The Dutch book industry and the distribution of books in a changing perspective", Presentation at the *32nd International Supply Chain Seminar*, Frankfurt, 2010, slides available at http://www.editeur.org/files/Events%20pdfs/Supply%20chain%202010/CB%20Supply%20Chain%20Seminar.pdf. A current presentation of the activities is at www.cb-logistics.nl/eng/e-books/.

metadata, of the DO. Relationships with a data aggregator like Europeana are more or less bilateral[71].

For a commercial partner, however, a whole chain (or more accurately, a network) of multilateral value-exchanging relationships pre-exist a new agreement with Europeana. For completion, content value is included in the description of the data value chain below, to emphasise that they are not seperable:

Data supply chain partner	Data value added	Data value received	Costs	Revenues
Content creator	Creation of intellectual property, upon which all data created is based	N/A (but could derive value from data on existing content for research)	Minimal or none - data probably created and maintained by publisher (unless creator is self-published)	Remuneration depends on contracts based on content, not data
Identifier management agency	Identifier creation, metadata assignment and management	N/A	Management of identifier infrastructure, metadata management	Registration fees
Content publisher	Identifier registration; intrinsic product description; sales rights data; marketing collateral; retail subject classifications; recommended retail price; discounts	Accurate product identification enables supply chain for competitive marketing and product sales	Cost of placing data in active marketing data services, or overhead of maintaining in-house data service	Indirect revenue from increased sales potential of products
Data aggregator / enrichment agency	Other data enrichments, editorial standardization of data, application of “house style”, data aggregation	Upstream data provides “raw materials” for selling data services	Data management and enrichment personnel and infrastructure; licensing of upstream data	Revenue from data service licensing to distributors and retailers

[71] The relevance of intermediary content and metadata aggregators for business models is discussed separately in the *D4.1 Addendum*.

Data supply chain partner	Data value added	Data value received	Costs	Revenues
Wholesaler / distributor	Supplier details; wholesale price	Upstream data enables management of ordering and sales of products	Data management personnel and infrastructure; licensing of upstream data	Revenue from wholesale of products
Retailer	Retail price	Upstream data enables marketing, management of ordering and sales of products	Data management personnel and infrastructure	Revenue from retail sales of products

The essential point here is that each stage of data creation, supply and use involves some value exchanged with partners in both up- and downstream directions. This value not only confers financial benefits but also adds some costs to the operation of the partners involved; for them to be (at minimum) self-supporting, therefore, revenues must be protected by contractual terms and conditions.

Any business model that builds on this existing value chain must (at least have the potential to) add to, rather than reduce, these revenues, since the profits from such activities are relatively small, and to reduce or remove them may make data creation, quality maintenance and enrichment unsustainable.

4.2 TECHNICAL REQUIREMENTS – LINKED HERITAGE AND EUROPEANA

A small number of technical requirements underlie the implementation of the basic business proposition. They are presented here in their relation to the costs and benefits listed above.

4.2.1 Europeana mandatory elements

Since its central asset is the metadata it holds, the minimum technical requirements for Europeana are concisely expressed by their ESE schema's mandatory elements. Some of these are expressed as alternatives (sets of elements from which *at least one* of the options must be provided), and the `dc:language` element is optional *unless* the `europeana:type` element has the value "TEXT" (this would provide, for example, the primary language of the product's text, in the case of commercial book products).

Europeana is currently in transition to a new schema, EDM, which contains substantially the same elements with some additions, primarily enabling distinctions and relations between the different entities described by the data (see diagram in section 3.3.6). With the exception of `edm:aggregatedCHO`, all of the mandatory elements for EDM map directly to elements from the ESE mandatory set, and are shown below in parentheses next to their equivalents.

EDM would have the advantages for commercial contributors of distinguishing the CHO, DO and preview descriptions, and thus enabling workable rights information, and sharing a central requirement with commercial product data: that of a universally unique, persistent identifier for the product itself.

Mandatory elements (ESE, with corresponding EDM element in brackets)	Use case(s) for commercial product data in Linked Heritage
europeana:provider (edm:provider)	Name or identifier of the provider of the product's data to Europeana directly – so in this case, "Linked Heritage". For data contributions in future this could include a standard identifier for the company, *e.g.* SAN, GLN, ISNI, of an intermediary aggregator, or the publisher themselves if contributing without intermediaries.
europeana:dataProvider (edm:dataProvider)	Name or identifier of the *original* provider of the product's data (either to Europeana directly, or to Linked Heritage as intermediary aggregator). As above, this could be a standard name identifier, and use of such identifiers is highly recommended so that contributed data can be traced back to its source if reused
europeana:rights (edm:rights)	General statement of rights held in the product (most likely copyright or similar creators' rights, not sales rights).
europeana:type (edm:type)	Simple classification of primary content type for Europeana's search portal faceting. Allowed values are TEXT, IMAGE, SOUND, VIDEO, 3D.
europeana:isShownBy and/or europeana:isShownAt (edm:isShownBy and/or edm:isShownAt)	HTTP URL links to either or both of: - Image representing the product; a book cover, DVD/film poster, album cover, thumbnail of photo – used by Europeana to generate preview images for its portal. - Web page with fuller details of the cultural heritage object (isShownAt) *i.e.* the publisher's or retailer's dedicated page where the product can be bought.
dc:language [when europeana:type="TEXT"]	The primary language of the product, if it is primarily text-based.
dc:title and/or dc:description	Minimal free-text information to identify the product and differentiate it from other similar products. The very open definition of dc:description can probably accommodate a very large selection of the wide range of descriptive text types used in the commercial sector, so it may be possible to select promotional texts intended for a broad public, and thus suitable for dissemination through Europeana's DEA.

Mandatory elements (ESE, with corresponding EDM element in brackets)	Use case(s) for commercial product data in Linked Heritage
dc:subject and/or dc:type and/or dc:coverage and/or dc:spatial	Minimal information to add value and interest to the product information, and enable contextual links to other items in Europeana. All of these fields can be reliably mapped to relevant fields in all four media industries which are normally populated from controlled value lists[72]. They will be mostly extrinsic metadata rather than intrinsic to the product, and so may represent value added at some mid-point in the supply chain. The dc:type element will probably represent a "product form" classification (balanced between content and carrier types) assigned at the early stages, resulting in the least value lost by disseminating under CC0.
edm:aggregatedCHO	Identifier of the source object itself (*i.e.* the product ID). Note that the corresponding ESE element, dc:identifier, is not mandatory, but the EDM element is listed here because of its commercial importance. Linked Heritage recommends that commercial data providers use at least the dc:identifier element containing a standard product identifier to ensure a minimum commercial potential in the data, assuming willingness to accept the risks inherent in CC0.

In the next section the ESE mandatory elements are shown in the context of the mandatory elements of the LIDO mediation schema used to aggregate data within the Linked Heritage project. Not all mandatory LIDO elements are mandatory in ESE, and vice versa.

The LIDO mandatory elements are extremely few, and all either identical with ESE/EDM requirements, or generated within the MINT system. For a LIDO record, the "title" of an object may in practice take the form of a short, encapsulating description, reflecting common practice in museums where a title as such may not exist.

The only other mandatory LIDO element worth considering here is the lido:recordID used to identify the set of metadata aggregated by Linked Heritage at a record (or data item) level.

As already described in Linked Heritage D4.2, LIDO is capable of fully expressing the precise semantics of commercial product data and could make an acceptable data aggregation schema. Its mandatory elements can be easily supplied by most commercial data contributors.

However, also as previously noted, the translation between rich LIDO data and basic ESE or even the more nuanced EDM is not simple, due to the numerous optional elements in the source schemas, and the lack of appropriate semantics – and display options – in the target schema(s).

[72] See Linked Heritage D4.2 (at: http://www.linkedheritage.org/getFile.php?id=394) and http://ddex.net/dd/ERN34-DSR40/DD/ddex_Class.html

Therefore, as with any aggregation schema, designing and maintaining these mappings will require investment in terms of expert staff and resources, especially as Europeana's schemas, API and portal display are under continual development.

N.B. in the table below, the LIDO XPATHs are shown truncated for brevity; the root XPATH is used as a section heading, shaded, at the top of each section. For more on XPATHs in schema mappings, see D4.2.

4.2.2 LIDO mandatory elements

LIDO ELEMENT	ESE ELEMENT	LIDO	ESE	COMMENTS
descriptiveMetadata/objectIdentificationWrap				
titleWrap/titleSet/appellationValue[@preferred="preferred" or “pref”]	dc:title	x	x	(dc:title mandatory if no dc:description; multiple LIDO sources for dc:description)
descriptiveMetadata/objectClassificationWrap				
/objectWorkTypeWrap/objectWorkType/term	dc:type	x	x	@type concatenated
/classificationWrap/classification (europeana)/[@type="europeana:type"]term	europeana:type		x	Predefined value list
/classificationWrap/classification[@type="language"/term]	dc:language		(x)	mandatory if europeana:type="TEXT"
descriptiveMetadata/objectRelationWrap				
(many LIDO elements map to dc:subject)	dc:subject		x	extentSubject concatenated
administrativeMetadata/recordWrap				
/recordID	dc:identifier	x		@type and "[Metadata]" concatenated
/recordSource/@type="europeana:dataProvider"]legalBodyName/appellationValue	europeana:dataProvider	x	x	
administrativeMetadata/resourceWrap				
/resourceSet/rightsResource/rightsType/term [@pref="preferred"]	europeana:rights		x	Predefined value list, see Europeana Rights Guidelines[73]

[73] See http://pro.europeana.eu/web/guest/available-rights-statements

4.2.3 MINT requirements

The MINT aggregation platform itself imposes technical requirements but also some minimal human intervention (once standard mappings are established) to process data ingestions and publication to Europeana:

Requirement	Comments
Manual upload	Once a Data Contribution is expressed in an XML format the upload process is simple and involves only a few steps.
FTP upload	FTP upload of XML Data Contributions is even simpler, but may raise concerns over security of the data.
Mapping support	Standard mappings of ONIX for Books 3.0.1 and 2.1 to LIDO were developed as part of Linked Heritage, as well as outline mappings of EIDR, IPTC and DDex that can be used to develop single standard mappings in MINT that, once produced, can be used many times by any Contributor.

The need to manage updates was addressed in D4.2 and now in section 4.4.8. Here it is simply worth noting that many of the technical elements here have been released as ready-made, "standard products" and certainly are in the public domain; Linked Heritage has published its LIDO mappings and the MINT software is available as an open source implementation so in principle Contributors can set up (and maintain) their own data feeds to Europeana in the Linked Heritage technical paradigm without direct, intensive reliance on Linked Heritage partners. This, of course, assumes the technical expertise to use MINT and follow the semantic mapping guidelines.

4.2.4 Minimal requirements of commercial sector data providers

Conversations with commercial players revealed concerns for the security of their data and previews, even within the Europeana.eu portal. Although discussed most comprehensively with image libraries, similar concerns, especially for product identifiers and retail links (which in combination would facilitate attempts at replication of service) were expressed by a number of other types of publishers and data services.

The most common technical requirements mentioned are listed here:

Requirement	Comments
Embedded metadata in images (including preview thumbnails)	Europeana acknowledges the potential to create effective orphan works if image previews are released without accompanying metadata, and that the best practice is to embed image data in the image file itself, persisting metadata from previous versions where available[74]

[74] See http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsPreviewMetadata

Requirement	Comments
Protection of thumbnail previews in Europeana portal from crawlers	Especially image libraries voiced concern about harvesting of image previews, which could have commercial value, especially if free of visible watermarks (the most obvious, cost-effective protective measure in commercial image search portals, but currently explicitly forbidden by Europeana). Steganographic (invisible) watermarks are not addressed[75].
Removal of data from Linked Heritage and Europeana	As discussed in section 4.4.8, a common requirement of commercial content and data providers is for their product data to updated within a reasonable number of days, or removed from circulation if no longer correct. Lack of this facility entails legal and/or commercial risks

Best practice in the commercial image sector is to persist metadata embedded in images as they are transmitted or altered, and a thumbnail version should contain identification, discovery and rights information inherited from its source[76]. Instead, Europeana proposes copying selected EDM fields into some XMP properties of the preview; the mapping they propose[77] is listed below:

XMP property in Thumbnail Preview	Constant value	Source field in Record EDM	Comments
dc:title[@xml:lang="x-default"]		dc:title	Useful for basic discovery and identification
dc:rights[@xml:lang="x-default"]		dc:rights	As with edm:rights, a plain text statement of rights ownership
dc:rights		edm:rights	As above
edm:dataProvider		edm:dataProvider	The name or identifier of the original supplier of the metadata
edm:provider		edm:provider	The name or identifier of the data aggregator; in this case, Linked Heritage
xmpRights:Marked	"True"		"True" indicates a "rights managed resource" as opposed to "public domain"[78]

[75] See for example, http://www.cs.bham.ac.uk/~mdr/teaching/modules03/security/students/SS5/Steganography.pdf or http://www.cs.usfca.edu/~brooks/S03classes/cs486/lectures/stego.ppt
[76] See D4.1, sections 5.3.5 and 6.6. (available at: http://www.linkedheritage.org/getFile.php?id=283) and also D4.2, section 12.2 (available at: http://www.linkedheritage.org/getFile.php?id=394)
[77] Taken from http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationPreviewMetadata
[78] See http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/cs6/XMPSpecificationPart1.pdf

XMP property in Thumbnail Preview	Constant value	Source field in Record EDM	Comments
xmpRights:WebStatement **cc:morePermissions**		edm:isShownAt	For Europeana, the URL of the product page
xmpMM:OriginalDocumentID		edm:object	The provider's identifier for the source image
cc:useGuidelines		[not applicable for commercial providers; "non-binding guidelines"[79] for use with public domain content]	This is the most unsuitable property for commercial content; see 3.3 on guidelines versus licence terms (commercial images can already use *e.g.* PLUS embedded licence terms[80])
cc:attributionName		dc:creator	As used in XMP, this is a straightforward mapping.
xmpMM:DerivedFrom		edm:shownBy [not recommended for use by commercial providers]	As used by Europeana, this would embed the URL of the source image at the provider's server enabling direct, automated access; something commercial providers avoid strongly

The above comments make clear that the EDM mapping is not ideal for commercial use; one-to-one persistence of all original metadata, except for IDs and other versioning information, is the best practice, which should be followed as closely as possible. It may be noted that Europeana's justification for omitting this function is that existing providers do not supply embedded metadata[81]; of course this would not be the case for commercial providers following best practice[82].

4.3 PROTOTYPE DATASETS WITHIN LINKED HERITAGE'S LIFECYCLE

As in the previous two deliverables of Linked Heritage WP4, the "research instrument" developed in order to fulfil the tasks specified by the DoW itself constitutes a valuable reference. These "instruments" are summarised below:

[79] See http://labs.creativecommons.org/tag/public-domain/ for context
[80] Full PLUS License Data Format here: http://www.useplus.com/useplus/license.asp
[81] See http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsPreviewMetadata#Need:forwardexistingmetadataembeddedinpreviews.
[82] For the sake of even-handedness, best practice is not always common practice. Numerous examples of commercial and public bodies, some very prominent, that remove embedded metadata when republishing images online can be found; hence the ongoing efforts of photography advocates to encourage a more consistent approach.

Deliverable	Research instrument	Benefits
D4.1	Descriptive metadata schema / identifier structure specification outline	In the context of commercial creative media product metadata and identifier standards, and representative cultural heritage (museum and library) standards, produced common framework for describing, comparing and evaluating standards' potential for interoperability in a rigorous, specific and detailed way
D4.2	Annotated XSLT and detailed narrative of semantic mapping of commercial sector data schemas	Provided full mapping for aggregation of ONIX for Books 3.0.1 and 2.1 records and clear guidance to produce mappings of EIDR, IPTC/XMP and DDex
D4.3	Commercial data contribution [authorisation] agreement	Illustrates real issues addressed in discussions with commercial sector partners; provides starting point for PPP

The prototype agreement produced for T4.2 forms the bridge between the technical mappings and specifications produced for D4.2 and specifying concrete legal forms for contributing data. It is based on the approach of working within the existing frameworks to produce a proof of concept which may not be ideal, though it is workable, before attempting more challenging agreements.

4.3.1 Prototype agreement

In D4.2 (section 20) guidelines for commercial data providers were given, describing the intended uses of the data within the project. The prototype agreement (in the appendix section 12) developed for this deliverable simply extrapolated clear contractual forms of these guidelines. In particular, the specific rights and duties included and excluded in the agreement were enumerated, and several open questions, not yet addressed by the existing Linked Heritage cooperation agreements, were closed, namely the ownership of the data once ingested to MINT, its status at the end of the project lifecycle and clear and precise statements of responsibility for ensuring security of the contributed data.

The new agreements are also simpler than the Linked Heritage cooperation agreements as they limit the actions of commercial partners purely to contributing data, rather than collaborating in a broader way with the project as a whole. On the other hand, the extended technical annex makes clear precisely which data elements are mandatory, and allows commercial contributors to specify exactly what, if anything, will be published to Europeana.

The new data contribution agreement also includes a contractual "wrapper" delineating the rights to receive, use and publish data given to EDItEUR as Linked Heritage partner from the duties to ensure data security assigned to the project coordinator as overseer of the technical and organisational infrastructure.

4.3.2 Paraphrase of the prototype data contribution agreement

Using the exact terms of the agreement, shown here in italics, a more readable paraphrase of its content is presented to aid understanding:

"The *Commercial Data Contribution Authorisation* agreement[83], between the Project *Coordinator* and *EDItEUR*, gives *EDItEUR* the rights to sign *Data Contribution Agreements* with *Contributors* and exercise the rights granted therein: to receive *Data Contributions* in *MINT* to test mappings to *LIDO*, pass *Publication Subsets* of *Data Contributions* designated as *Prototype Data to Europeana*, and sign the *DEA* for those *Publication Subsets*; it commits the *Coordinator* to ensure that no part of any *Data Contribution*, except *Publication Subsets* of *Prototype Data*, is released outside of *MINT*, while ensuring the *Data Contributor* retains all rights in the *Data Contribution* and can access it at any time throughout the *Linked Heritage* project (at the end of which, the *Coordinator* guarantees that all *Data Contributions* will be deleted from *MINT*).

Data Contribution Agreements enabled by the *Commercial Data Contribution Authorisation* give *EDItEUR* the rights to receive *Data Contributions* in *MINT* to test mappings to *LIDO*, pass *Publication Subsets* of *Data Contributions* designated as *Prototype Data* to *Europeana*, and sign the *DEA* for those *Publication Subsets* (both actions involving *Prototype Data* subject to written final confirmation from the *Contributor*), as well as the duties to ensure that no part of any *Data Contribution*, except *Publication Subsets* of *Prototype Data*, is released outside of *MINT*, while ensuring the *Data Contributor* retains all rights in the *Data Contribution* and can access it at any time throughout the *Linked Heritage* project (these duties are fulfilled by *EDItEUR* by signing the *Commercial Data Contribution Authorisation* agreement with the *Coordinator*).

The *Contributor* has the rights to retain full ownership of all *Data Contributions*, request the removal of *Data Contributions* from *MINT*, specify whether a *Data Contribution* is *Test* or *Prototype Data*, and for the latter specify the *Publication Subset*, including whether to include URLs of *Thumbnail Previews* in *Europeana* from data released under CC0, and choose whether or not to give confirmation of permission to pass *Publication Subsets* to *Europeana*. *Contributors* have the duty to provide the *Data Contribution* specified in the technical annex, and to warrant that all data contained in the *Publication Subset* of any and each *Data Contribution* designated as *Prototype Data* has been cleared for release by *Europeana* under the terms of the *DEA*, principally the Creative Commons CC0 terms."

These terms represent the maximum security and assurance for commercial sector contributors possible under the current Linked Heritage and Europeana frameworks.

4.3.3 Terms and conditions

The terms (*in italics*) of the prototype data contribution agreement summarised above are included in the glossary in section 9, also *in italics.* These terms will also be used throughout section 4.3.5 to make the connection between the broader discussion of the challenges involved in large-scale aggregation and the agreement illustrating what is possible now.

4.3.4 The missing link – agreements with Europeana

Some elements of the agreement that were requested by commercial contacts could not be included in the agreements for Prototype Data because they are out of the control of EDItEUR and the Coordinator. The essential elements were:

- *Frequency of updates*
  Europeana's aggregation systems update monthly, but commercial sector data must be updated as rapidly as reasonably possible, for example, within one or two business

[83] This agreement was created solely to make the process of signing Data Contribution agreements simpler, by enabling EDItEUR to – in effect – act on behalf of the Project Coordinator. The two agreements could easily be combined into a single agreement between the Project Coordinator (or a legal body acting on behalf of the project) and the Commercial Data Suppliers.

days[84] in case of change of product availability or a legal issue with the product or the metadata.

- *Selection, display and later modification of Publication Subset*
  The options provided in the prototype agreement presented here follow the mandatory fields of EDM/ESE. Ideally, a commercial provider would retain full control over selection of data elements and some control over their arrangement on screen and presentational context.
- *Protection of rights in images*
  All images used in commercial sector metadata are covered by copyright (as in fact are any type of images created by human agency) and will normally be licensed for specific uses, such as creation of a book cover and use of the thumbnail or larger cover images in online retail platforms. The contract governing use of the cover images generally will rule out other uses of the image, and indeed may not have been exclusive (*i.e.* the same image may be licensed for other uses and/or by other licensees). This is clear from the terms of many book data services, even where the textual data, as in the example below, is freely licensed:

  > "I understand that the material I download from Datashop is in copyright and is made available for the purpose of promoting/publicising the title(s) to which it relates, and/or for personal use, and that it may not be sold or used for other purposes."[85]

  The need for such warnings is clear – book publishers often do not license broad enough rights in the component parts of book covers (eg photographs included on the cover) that would allow them to sub-license the completed book cover for other purposes.
  Although rights information can be included in the ESE/EDM metadata presented in the Europeana portal, API and linked data dumps, it is not strongly linked to the image according to the best practice of the photography industry, in that
  a) The ESE/EDM elements are not protected by *sui generis* database right due to CC0 and therefore could be easily removed (in fact this will be encouraged by CC0);
  b) Rights information is not yet embedded in the thumbnail file itself[86];
  c) There is no technical protection to prevent stripping of thumbnail images from the portal or via the API, even if URLs for them are not included in the linked open data release.
- *Removal of data from Europeana (linked data)*
  Finally, as described already in footnote 84 and section 4.4.8, there is no mechanism yet for rapid takedowns from the Europeana portal – but more crucially, once data has been universally published under CC0, there is no legal obligation for any sort of takedown, either by Europeana or any other third part data reuser. "Community norms" and unenforceable guidelines do not provide any guarantee in this case, but legal and commercial risks from inaccurate data are known from commercial players' experiences.

[84] Book industry best practice, for example, indicates a maximum of 10 business days for routine updates: http://www.bic.org.uk/files/pdfs/110721recipients%20best%20practice%20final.pdf – note that updates such as emergency "takedowns" (as in section 4.4.8) need to happen much more rapidly.
[85] Pop-up terms and conditions acknowledgement from Cambridge University Press DataShop (see http://datashop.cambridge.org/) – note that these terms are only displayed on attempting to download cover image files.
[86] Even though this option has been considered by Europeana: http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsPreviewMetadata#Need:forwardexistingmetadataembeddedinpreviews

Since CC0 avoids warranties of title, a potential reuser can never be sure that the data under CC0 is genuinely freely usable, and there is every possibility of takedown demands from other rightsholders. To avoid issues like this, **the terms of the prototype agreement (Appendix Section** 12**) provide warranties of title**, which can guarantee that no other party has a claim on the data.

4.3.5 Data owners and data service providers

In the heritage sector it is relatively likely that an organisation characterised as a data provider will also be the creator and owner of the rights in the data provided. No "partners" are, in principle, required to complete the "supply" of the data describing unique items when the curator, archivist or librarian has direct access to every aspect of the item's presentation to the public. This situation may change if "User Generated Content" (UGC) becomes commonplace; and of course we have generalised here – many heritage organisations actually do "outsource" (see 3).

In the context of Europeana, the situation is often somewhat complicated by the use of aggregators to deliver data services, such as the MINT instance used in Linked Heritage to enable data storage, mapping, enrichment and publication support to a large number of data contributors, but this is not strictly essential. In fact, Linked Heritage aims to facilitate standardisation and aggration of heritage institutions' data – which the institutions could in theory do themselves[87]. The aim of the various Europeana Network projects is not to provide services "on the market" but simply easing the resource burdens of administration and technical collaboration efforts.

However, in the commercial context, at any reasonably developed scale, the situation is far more complex: data creation, storage and distribution services already exist at various levels of partnership with the publishers of products. The small selection of examples below is for illustrative purposes only.

- Many metadata service providers offer IT platforms to publishers enabling a basic level of data production;
  - BooksoniX
  - Stison
  - Focus on Publishing
  - ONIXSuite
  - Schilling
- Commercial data aggregation services act as intermediaries between publishers and retailers, wholesalers and other logistics organizations which also deal in the associated metadata;
  - Nielsen BookData
  - BDS
  - Bowker
  - IE Online
  - DILVE
- In the e-book and e-music worlds, digital asset distributors trade both the digital products and the associated metadata.
  - 7Digital
  - ePubDirect
  - Ditto Music

[87] For example, both the full LIDO data format specifications (www.lido-schema.org/) and the MINT aggregator itself (http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mint) are freely available on an open source basis.

- Consolidated Independent
- Ingram Content Group
- The Orchard
- TuneCore

Therefore the technical provider of a commercial dataset will often be a distinct organisation from the data owner, and the content and metadata value chains intertwine in a complex manner.

4.4 DIMENSIONS OF BUSINESS MODELS

Moving beyond the minimum requirements for an agreement within the current framework, the detailed specifications outlining the "why", "what" and "how" of specific agreements were explored.

4.4.1 Business Models for Content Aggregation – "the partnership"

Initially it should be stressed that the type of business models considered here will be very different from those previously envisaged in the context of Europeana and its associated projects (*e.g.* in European Commission, 2011, 2003 or Savenije, B. and Beunen, A., 2012). A brief comparison of partnerships to achieve heritage digitisation ("content creation") and those to enable commercial metadata exchange ("content aggregation") should highlight this:

Aspects of partnership	Digitisation	Metadata exchange
Output corpus definition	Finite number of items (a fixed corpus of content is identified as the target of digitisation)	Potentially open-ended set of (data) items (new content items are published all the time and generate new metadata items)
Potential to delineate public domain and exclusive sets of content	"Moving wall" or embargo ("diachronic") on release of some content items (often the embargo period runs from the digitisation date, but it could also be based on the creation dates of the original analogue materials)	Separation of each item into browsable and reusable subsets of data elements, possibly at API or RDF dump level (or each dataset into free and subscribed subsets, possibly available through Web links to a separate portal) ("synchronic")
Increase in availability of output to the public	Significant increase through new online availability	Data is already available, often through multiple online outlets
Integrating maintenance of access into business as usual	Output is part of public access mission of institution	Output is an additional data channel needing extra maintenance that must be justified

The differences in these types of partnership in many ways parallel those identified in D4.2 (Appendix 2) between data modelling of heritage objects and commercial products and just as those technical differences meant a new type of technical solution, so here a new legal and financial model of partnership will most likely be required.

Above all, the primary difference between the types of "content" considered here (apart from the obviously far smaller *quantity* of information in the metadata exchange case) is that for commercial data, *currency* strongly correlates to *value.* An embargo on commercial product data based on the currency of the products would reduce the data's value to its end users, the potential to offer meaningful exclusivity rights to the content (data service) producer, and thus the whole partnership's viability.

In the case of a heritage partnership, this may not be a strongly pronounced difference, since the market for cultural heritage media products probably remains closer to the development of its subject matter which is of enduring interest. If there were some link to specific time-bound events, such as special exhibitions, new discoveries or prominent anniversaries, the embargo model might, however, apply.

4.4.2 Defining the partnership – contracting partners

Any partnership with the commercial sector's involvement should be conducted on the basis of *robust contractual agreements* between clearly identified partners. In particular these should make clear:

1. Which rights are retained by the contributor in the data contributed;
2. Which rights are transferred to the recipient of the data, and more importantly, which rights are *not* transferred, in terms of
    a. Use by the recipient, and
    b. Uses that may be granted by the recipient to third parties;
3. Who is responsible for ensuring *security* of the data exchanged;
4. The intended scope and particularly the *limits* of the rights granted by the contract, even to the point of specifying what is and is not implied by each particular right granted.

The *benefits* of this approach go beyond what could be achieved with a simple memorandum of understanding, for several reasons:

a) A detailed contract can clarify the rights and duties of each party and sets the collaboration on a transparent basis;
b) Beyond this, inclusion of standard terms and conditions can exclude known legal risks, some of them unique (at least in their severity) to the commercial sector, involving parties not currently contracted to the commercial partner;
c) Another vital requirement for the commercial partner is to differentiate the new agreement from existing contracts with other commercial partners to ensure that there is no conflict of interest.

This approach highlights the requirement for any agreement to be concluded between identifiable, known and trusted partners. In the current context of specifying requirements for prototype data contributions, the partners in any agreement are clear:

1. The data contributor (a commercial organisation outside the Linked Heritage consortium);
2. (Possibly) EDItEUR (acting as a data recipient and trusted partner for the commercial organisation);
3. The Linked Heritage Project Coordinator (acting as the responsible partner of EDItEUR as data contributor to Europeana, principally to ensure security of the data).

Because the Linked Heritage Project Coordinator is able to assume duties beyond the lifetime of the Linked heritage project, there is a measure of security and sustainability even in this arrangement, although transfer of these responsibilities to a long-lived legal entity such as the Michael Culture Association is clearly preferable. However, for any large-scale or longer term data contribution agreement, legal partners would have to be found who could assume the rights

and duties, here carried out by EDItEUR, of receiving and maintaining the data contribution, as well as assisting the data contributor in the long term. This need will be addressed in 4.6.1.

4.4.3 Commercial sector product data – "the offering"

It may seem counter-intuitive to follow a discussion of contracts with a section revisiting the topic of the basic nature of commercial product metadata, since that topic has been extensively discussed both in D4.1 and D4.2. However, the aspect to be highlighted here is that *at every stage of the product supply chain – and by implication, the metadata supply chain – product metadata is essentially used to support commercial transactions* which take place within contractual frameworks like those described above. Most prominent metadata standards support the initial phases of commerce, making an offer to trade, but other important standards, based on the same product information, support the completion of some kind of sale, or the later stages of sales reporting and facilitating or reporting secondary payments such as royalties.

A simplified comparison of standard messages from the media sectors described in D4.1 and D4.2, and their purpose within the supply chain can illustrate this:

Media sector	***Uses***			
	Identify products and their content for use in transactions	**Describe product (and/or content) as object of transaction**	**Define transactions**	**Report transactions and pay third parties**
Recorded music	ISWC, ISRC and GRiD reference metadata messages	**DDex ENRM** suite	DDex ENRM Business Profiles, Musical Work Licensing Message suite	DDex Digital Sales Reporting profile
Books	ONIX for ISBN, ONIX for ISTC	**ONIX for Books**	EDItX Consumer Direct Fulfilment	EDItX Sales Report
Film and TV	**EIDR**	**EIDR**		
Photography		**IPTC/XMP**	**IPTC/XMP [PLUS fields]**	

The product description/identification schemas addressed in Linked Heritage WP4 are highlighted above, showing that even those formats closely resembling repository metadata have an implicit or even explicit use in making a commercial offering to one or more supply chain partners or to the public at which they are targeted.

The close relationship between, and ultimately, identity of "descriptive" and "rights" data becomes especially clear in the case of commercial images, where generally what is on offer are the rights to use the image file or versions thereof under very closely specified terms and conditions; this view of "product as set of rights" can be generalised to other media sectors, and in the digital world, it becomes increasingly obvious that this is the correct view as encodings, formats and

technical protection play a role alongside the ease with which copies or variants of works are made (see 3.5).

4.4.4 Defining the offering – whose product?

As already noted in 1.2.1, and throughout D4.2, commercial products are represented as *classes of items*, as opposed to *unique heritage items* and this difference introduces another important problem: not only are various data supply partners used in many cases, but more than one *retailer* can offer the same product at the point of sale.

This means that, in analogy to Europeana aggregations, there can be several candidate URLs for the edm:shownAt field, where in heritage institutions one would expect to find only one clear source for this Web page. Unless one considers that "the most obvious default for the "repository" of a product type is the publisher"[88] and decides to use only data containing a link back to the Web site of the current publisher (or other releasing organisation) of each product, there is no clear method for selecting retail links to map into edm:shownAt.

It is acknowledged that "publisher-as-retailer" is a big assumption, and an over-simplification of the real situation; simply put, dedicated retailers are able to offer economies of scale when it comes to making sales not available to publishers who have other core business needs.

This simplification may have commercial consequences; data service providers may wish to retain commercial neutrality and avoid including *any* retail links in their service (or more problematically, may wish to include multiple links to all possible outlets); on the other hand, there may be a legal risk relating to changes of availability (see 4.4.8) or even to the underlying publication rights (see 4.5.2).

4.4.5 Data as a core business asset – "the service"

Only in the light of data messages intended in this overtly commercial way can the concerns of commercial data contributors be understood. Because products (and sets of products) have a commercial value, the data providing information about them and access to their suppliers has commercial value for those in the supply chain who wish to exploit the products. As an illustration, one senior publishing director described the problem:

> "Regarding the risk of undermining commercial models, here's the gist:
>
> A number of organisations operate a (for-profit) link resolving/indexing service for libraries who pay them a subscription;
>
> - These businesses supply a hub service to libraries which consolidates all sources of e-book purchases to a single hub from where library members can search/browse across all collections (irrespective of where purchased), and this greatly reduces the complexity to library users of engaging with a number of platforms separately;
> - There are other businesses that license or buy our metadata for other reasons;
> - I believe that a Europeana repository of book metadata which would offer that data for free as CC0 might undermine our current commercial deals.

> It's a risk which I and my colleagues need to ensure is either unfounded, or acceptable. Four years ago our view of metadata was that it had no intrinsic value as such, and that 'letting it go' had a promotional benefit to the publisher. Things have changed in the intervening time where we realise that people can (and do) make money from our metadata and therefore it's reasonable to expect to realise some of that revenue ourselves."

[88] See Linked Heritage, D4.2, section 17.2.2. at http://www.linkedheritage.org/getFile.php?id=394

– Mark Majurey, Commercial Director, Digital Publishing at Taylor & Francis Group.

Clearly in such a situation, an additional risk from open publication of raw product data would be that of losing a high-quality, authenticated service provided by the source of the product described by the data. The argument for the growth of open (usually public-sector) data, that the value of the combined dataset will increase with increased use, simply cannot apply to provision of commercial product data:

> "In the case of open knowledge the feedback arises because, with viral licensing, each user of the knowledge pool becomes a contributor back to the pool. As the pool grows it is ever more attractive to new users so they use (and contribute) to it rather than to any competing closed set of knowledge."[89]

In the case of commercial product data, **the ultimate value of pooled datasets is more likely to fall as the number of errors in the dataset will increase with reuse and lack of updates from the original provider.** Because maintaining, updating and (where necessary) correcting the data takes time and funds, data reusers who will most likely not receive any income from the data they contribute will be unable to maintain a high quality dataset, while any drop in income from loss of paying customers who choose to use the free (reused) datasets will discourage the original provider from maintaining the (costly!) authoritative data service. See case 4.5.10 for illustration.

Essentially, this is an argument that the value of a data pool lies not in the volume of the pool, but in the accuracy of its contents – and that accuracy requires management effort to maintain.

A data service of this type has three main aspects; exposure, coverage and currency. These will now be discussed.

4.4.6 Defining the service – exposure

Considered as a closed, coherent system, a data service has a definite end-point, the ultimate users of the data it provides, for whom the data's content and context has a value. As well as being the fundamental justification for providing the service, aspects of this "audience" can offer a quantitative measure of the appropriate financial value of the service.

4.4.7 Defining the service – coverage

As can be seen in the above illustrative example, the ability to offer central coverage of a publisher's entire output is seen as bearing considerable commercial value. Coverage of the entire output is crucial:

1. A trusted, complete source means no time, effort and funds need be expended collating the information in-house;
2. Access to the full dataset means that each search result has more value in the context of all available records; primarily, its greater relevance is guaranteed by the fact that the dataset has in effect already been edited.

One aspect of coverage is far simpler to address in the context of Linked Heritage and Europeana's existing framework. A direct approach would simply be to receive Records for products that have been classified (either "intrinsically" by the releaser of the product, or "extrinsically" by Linked Heritage) as relevant to "customers" who might find the data through

[89] From an anonymous post at the OKFN blog; see http://blog.okfn.org/2006/11/06/open-knowledge-drives-out-closed/

Europeana. The aim would be to select a subset of the Contributor's datasets that is specific – and small – enough not to replicate any commercially-contracted services.

4.4.8 Defining the service – currency

Currency of data is essential as noted above, because of the perceived value in the ability to offer a single access point to a known set of products and services; updates in those products and services must be reflected by updates in the number and content of data Records available in the data service.

Currency also has implications for each Record's value since factors such as price and availability can affect direct commercial demand for the Product described by the Record. In fact, in the case that data Records are not updated to reflect a rise in price, or cessation of availability of the Product from a given supplier, the data may even reflect a negative value. It will give rise to reputational risk due to loss of trust in the data service (this is the "Brand Name" issue described in 3.3.3)[90].

The opposite effect applies in this case; it is in the interest of the data provider to contribute current data to any partner involved. For those partners receiving data, implicitly or explicitly, there is a duty to accurately represent the rights held over all products described, and their current sales status. Managing updates between a known set of partners, with no "anonymous" third-party reuse, is far more efficient that releasing product data universally for reuse, because considerable extra effort and resources would be expended to maintain comprehensive provenance data for each update (rather than leaving it implicitly within the technical running of service to each known partner).

Unfortunately, in the current Europeana framework, this is the most difficult aspect to control, both technically and because of Europeana's current policy:

1. Europeana's dataset updates run monthly[91], but, especially for digital products, changes of wholesale and retail availability and price change much faster and are expected to be communicated as quickly as reasonably possible. Common guidance for book publishers and retailers, for example, suggests updates should be actioned within two working days, and certainly no more than five days [92];
2. The CC0 rights waiver means that there is no obligation for all reusers to republish only the most recent updates, and in any case, it removes the need to understand and agree to use the message exchange mechanisms for *e.g.* deduplicating contradictory messages by preferring more authoritative message sources.

The need to manage updates according to best practice was addressed in the recommendations of Linked Heritage D4.2, sections 14.2.7 and 14.2.9; since it cannot be further explored here it will be left for final recommendations in the last chapters of this report.

For Linked Heritage there is one clear outcome – on project timescale measured in months, there is no scope for including any updates in the prototype data contribution agreement.

[90] In the case of picture licensing, a mostly business-to-business activity with a high per-product transaction frequency, there is also a potential legal and reputational risk if ownership of licensing rights (for example, of a large, unique collection of original images) changes hands; if the data describing this is not up to date it may simply be entirely wrong and could be construed as negligently misleading.
[91] Note that for rights transfers, for example in picture licensing as above (footnote 90), monthly updates *may* be acceptable. Shorter timescales are mainly a requirement for retail sale.
[92] See http://www.bic.org.uk/files/pdfs/110721recipients%20best%20practice%20final.pdf for an example

The remaining sections below will explore in detail how the coverage of Data Contributions can be specified in order to enable commercial sector organisations to respect existing commercial agreements while still providing cultural value to Linked Heritage and Europeana.

4.4.9 Quantity of data contribution – "the dataset"

This dimension measures the number of product records to be delivered to Linked Heritage and Europeana, either as an absolute number of records, or a percentage of a data provider's repertoire (title list, catalogue), or as both. This number of data records will be referred to as "the dataset" (to be contributed).

Although in the context of Europeana, Linked Heritage and related projects, terms like "number of objects" are often used, here we will refer simply to the number of product records. Even when there are easily identified sets of related products such as hardback and paperback versions of the same text; DVD and BluRay of a single movie or TV series; CD or MP3 downloads of a musical track; high or lower resolution versions of one image – all these alternative versions would clearly mean extra product records).

4.4.10 Defining the dataset – Europeana's stated aims

The Europeana Content Development Strategy[93] provides several clear categories to begin a subject classification definition of the dataset:

- "Digital cultural content"; defined by
  - Sourced from a cultural heritage institution;
  - Availability online in an accessible, digital form;
  - Perhaps belonging to a specific collection of particular interest, either
    - based on partners' own collection development strategies;
    - national collection development strategies;
    - domain (*e.g.* museums, galleries, film archives, sound archives…);
  - "European thematic collections not elsewhere covered" (*e.g.* Art Nouveau; First World War; European identity; European sport heritage; Napoleonic wars…)
  - Specific types of objects spread across different collections in Europe (*e.g.* Erasmus' letters; images of the goddess Athena…);
- Specific countries' digital cultural content – primarily those as yet under-represented;
- Specific content types – especially, in Europeana's classification, video and sound;
- Specific users' interests, including
  - academic researchers;
  - heritage professionals;
  - teachers and ***students***;
  - enthusiasts and amateurs such as
    - travellers/***tourists***;
    - cultural festival-goers;
    - local or family historians;
    - hobbyists working in specific media and genres such as film, photography, classical or early music, literature or poetry.
- Other categories compounded from five aspects as in the table overleaf.

The two categories highlighted above, secondary school students and cultural tourists, are singled out for special mention as markets for the redistribution of Europeana's content

[93] Avalable from http://pro.europeana.eu/documents/866205/0/EV1-AF-ContentDevStrategy.pdf

(metadata and previews), through projects such as Europeana Creative[94]. This is addressed in *D4.1 Addendum*.

In an update to this foundational document, Europeana (2012) confirmed the above areas of positive interest and the categories listed overleaf where more content is especially required for balance (and to meet existing EC targets). Also in the 2012 update was a detailed discussion of "masterpieces" from each EC member state that must be included in Europeana, reporting a piece of research, currently in progress, enabling Europeana's partners to identify the masterpieces of their own country and collections, and categorising them according to Europeana's media types (TEXT, IMAGE, VIDEO, AUDIO, 3D) and type of public access. As the masterpieces research progresses, datasets could be selected by their relevance to these specific objects and works, resulting in a highly targeted method of content selection, and a clear method for fostering cross-sector dialogue.

As well as selecting content *about* masterpieces, products *containing* modern masterpieces (defined more loosely than by inclusion in Europeana's own categorised list) themselves could be selected, for instance by collating lists of literary and artistic prize winners and their works. This content selection could also be reflected in business cases that call for external grant funding or ongoing subsidy, as creation and dissemination of metadata for such prizes, as well, perhaps, as targetted translations might be funded as part of the publicity budget of the promoting organisation.

[94] See http://pro.europeana.eu/web/europeana-creative for more details; proposed areas of work are currently "History Education, Natural History Education, Tourism, Social Networks, and Design".

Table 6 - weakly represented content categories in Europeana (as of December 2010)

	TEXT	IMAGE	VIDEO	SOUND
Subject	Musical scores and lyrics Performing arts and film Political documents	Technical Archaeology Medical Biology Economy	Silent film News Interviews	Jazz Contemporary music Classical music Wildlife sounds Ethnographic recordings
Period	Prehistory – 14th Century Contemporary	Prehistory – 14th Century	Prehistory – 19th Century	Prehistory – 19th Century
Language	Bulgarian, Icelandic, Latvian, Lithuanian, Maltese, Portuguese, Romanian	Bulgarian, Czech, Danish, Estonian, Finnish, Greek, Hungarian, Irish, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Slovak	Bulgarian, Estonian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Slovak, Slovene	Bulgarian, Czech, Danish, Dutch, Estonian, Greek, Hungarian, Latvian, Lithuanian, Maltese, Portuguese, Romanian, Slovak, Spanish
Country	Belgium, Bulgaria, Cyprus, Estonia, Iceland, Latvia, Lithuania, Luxembourg, Poland, Romania, Slovakia, Sweden, Switzerland	Cyprus, Czech Republic, Finland, Hungary, Latvia, Lithuania, Luxembourg, Macedonia, Norway, Poland, Portugal, Serbia, Estonia, Iceland, Slovakia, United Kingdom	Belgium, Bulgaria, Cyprus, Estonia, Finland, Greece, Hungary, Iceland, Latvia, Lithuania, Luxembourg, Norway, Poland, Romania, Slovakia, Slovenia, Sweden	Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Luxembourg, Poland, Portugal, Romania, Slovakia, Spain, Sweden

Despite its stated interest in *serving* the academic audience it is clear that Europeana does not intend to incorporate published content *produced* by that audience. Its policy excludes:

- "High level scientific content post 1950";
- Datasets, research papers and theses.

Although some of this content may be highly commercial, scientific journals and primary data have been explicitly excluded from WP4's scope[95].

[95] See D4.1, section 2.3.2, available at: http://www.linkedheritage.org/getFile.php?id=283

4.4.11 Defining the dataset – product form

Given the stated aim of Europeana to make available *digital* European culture, it is reasonable to focus on *digital* products (either born digital, or derived from some physical product), which are available online, for inclusion in the dataset.

Table 7 - conditions for preferred product forms in each media sector

Media sector	Standard	Data element	Preferred value(s)
Books	ONIX for Books 3.0	ONIXMessage/Product/DescriptiveDetail/ProductForm	AJ, EB, EC, ED, LC
	ONIX for Books 2.1	ONIXMessage/Product/ProductForm	AJ, DG
Film and TV	EIDR	FullMetadata/ExtraObjectMetadata/ PackagingInfo/PackagingClass	Streaming (Web), Download (Web)
Music	DDex	n/a	Product form details from retailer data
Photo	IPTC/XMP	n/a	Only used for digital files; further product form details from retailer data

Apart from this, concentrating on digital products means that more records are likely to have URLs, as born-digital products and their data will be prepared with the online retail environment in view; it will encourage use of the latest version of standards (*e.g.* the newer ONIX for Books 3.0 was significantly updated to describe e-books); and there may be separate databases for physical and digital products in many data services.

4.4.12 Defining the dataset – subject classifications

The BIC and BISAC subject classifications for book retail form a practical basis to define datasets, since these (with local variations of BIC in the EU[96]) are used widely in the commercial book trade in many European countries to describe the subject matter of in-commerce books. Even though the BISAC codes originate in the North American book trade, they are still widely *used* in the EU, for example, in order to sell through American retail sites, and their inclusion may help to capture books *published* in the United States *about* European topics.

Subject classifications for commercial book trade, and indeed classifications by *e.g.* genre or style in commercial film and music, tend to be relatively "flat" or simplified compared to those employed in the library and wider heritage world[97]. This is primarily because of such practicalities as:

[96] For example, CCE in Italy (available at: http://www.ie-online.it/CCE2_2.0.pdf) and a similar BIC localisation in Spain. In addition, a new initiative to create a single harmonisation across the USA, Canada and EU (plus other countries) has recently begun; see http://www.panthema.org/

[97] Of course, publishers, retailers and wholesalers that supply directly to libraries may suggest library classifications as an added value service, as in the case study presented in 4.5.5.

a) the operational need to keep data creation simple, cost-effective and quick to respond to changes in markets;
b) commercial considerations, such as
    a. the primary aim of general book retail ("trade") to reach a broad market of buyers, whereas libraries aim to reach a more specific "audience";
    b. the relatively small number of *current* publications on sale in any given subject, compared to the tendency of research and education libraries to collect larger numbers of publications on narrower, related topics over time; thus in-commerce publications can be allocated a more general subject heading on the understanding they will be differentiated from similar products by other factors (*e.g.* author, publisher, publication date, prizes, reviews, relation to other current products, sales levels…).

Hence the commercial schemes considered here either have a small hierarchy of terms (between two levels and eight levels for BIC; up to three for BISAC).The need for expansion of these hierarchies in *some* highly commercial areas is recognised, but has not yet led to adoption of a completely modular and extensible approach as in some libraries.

BIC and BISAC do, however, each have a set of "qualifier" terms similar to those used in library classifications like Dewey Decimal or Library of Congress that modify the scope of the subject by geographical or temporal coverage, or intended audience. These qualifiers are far less complex than the subject "facets" used by some libraries, for example in UDC[98]. Using geographic and time period qualifiers in some way to control the relevance of the products' content to Europeana's stated aims is highly desirable, both in terms of the cultural richness added to Europeana's database, and in terms of presenting a viable business case to the data contributor, who may be wary of simply uploading their entire dataset.

BISAC and BIC geographical qualifiers are,rarely used in practice (at least in comparison to the main BIC and BISAC subject classifications); however, if present, they can be used to exclude anything other than the codes for Europe and its constituent regions, allowing also records with no geographic qualifier. A proxy for geographic subject qualifiers may well be used; for example, if it is possible to select publishers known to produce titles only European topics, or if relevant ascribed or bibliographic collections[99] can be identified.

Defining the dataset provided by each commercial partner by subject classification will certainly reduce the total number of product records contributed, but it should also have the effect of increasing the overall quality of all commercial datasets taken together, since the relevance of the full set of products to the end user will be higher (on average) and thus finding relevant products will be statistically much easier.

The quality of the dataset as a starting point for linked data applications should also be improved because it will be based on known, structured semantic criteria, similar to those already promoted by the W3C (*e.g.* SKOS).

4.4.13 Quality of data contribution – "the template"

The record "template" refers to the *selection of data elements actually provided* to Europeana – the Publication Subset, in terms of the Prototype Data Agreement (see Appendix section 12). It is specified in the Technical Annex to the agreement, whose wording ensures that it will always comprise:

a) At least the mandatory ESE/EDM elements to satisfy Europeana's requirements;

[98] See http://www.udcc.org/index.php/site/page?view=about_structure
[99] See http://www.editeur.org/files/ONIX%203/ONIX_Books_Sets_and_Series_3.pdf

b) Any optional extra elements the contributor wishes to provide.

4.4.14 Defining the template – included information

The specific fields to include will depend on their business value to the provider; as in 4.1, each partner in the supply chain creates and maintains data fields that are sustainable in terms of cost to produce, value added and revenues from their clients.

Levels of "template value" could be estimated by comparing the ESE/EDM and LIDO requirements with minimal dataset associated with specific basic services in the commercial services, principally:

- Identifier registries, *e.g.*
  - ISBN, ISTC for books;
  - ISRC, GRID, ISWC for music;
  - EIDR for film and TV;
- Commercial-to-heritage data supplies, *e.g.* cataloguing-in-publication (CIP) programmes[100].

These templates provide an absolute minimum set for identification, disambiguation and approximate categorisation, in the case of CIP, also with an indicative price for budget-based selection, or, in the case of EIDR and ISBN-A, actionable identifier services. Though they have minimal value for the *data* supply chain, and relatively little *cultural value* in terms of rich descriptions, they may have the advantage of *being mostly intrinsic metadata* and supporting the *content supply chain*, so their creators may be more willing to share them openly.

4.4.15 Defining the template – granularity of data

With most commercial data formats it is possible to create data with a very high degree of granularity; for example, the ONIX for Books 3.0.1 specification defines almost 1000 possible unique data elements; in a typical Product record, perhaps 150 – 400 lines of XML will be used, representing a comparable number of data elements. Even a more compact format like DDEX may run to around 500 lines due to the complexity and richness of its content.

Converting such data to ESE or even EDM will thus necessarily involve loss of detail, either in absolute terms (less included information because there are fewer data elements) or in terms of the specificity of the resulting information (*e.g.* a "date of first publication of the work" embodied in a modern edition of a classic book in ONIX could simply become "a date associated with the resource" in ESE). Because ESE/EDM do not (yet) allow for flexible, configurable semantic relationships between data elements as LIDO does, the conversion is always from more to less detail; even if the data content from several distinct granular data fields are still included by concatenating them, there can no longer be a separate data type associated with each part of the concatenation, except by including human-readable separators and requiring reusers to perform significant ad-hoc development if they wish to make the data machine-readable. Unlike the domain-specific minimal datasets listed in 4.4.14, Dublin Core-based formats (that do not make use of highly-specific application profiles) are not semantically expressive enough to transmit even this basic information without loss or increased ambiguity.

Ultimately commercial data providers will need to make decision on the utility of the "lower resolution" data for providing enough discoverability and access to their product to make some kind of commercial incentive to sharing the data, and incurring the costs of maintaining a feed.

[100] For example, at the Library of Congress (http://www.loc.gov/publish/cip/techinfo/corematerials.html) or the British Library (http://www.bl.uk/bibliographic/cip.html and http://www.bl.uk/bibliographic/pdfs/cip.pdf).

4.5 CASE STUDIES – COMMERCIAL OFFERINGS & DEGREES OF OPENNESS

Here we present 13 business model case studies documenting the final outcome of conversations with book publishers, data service providers in the book trade and image libraries. Although contacts in the recorded music and film & TV industries assisted this work with some extremely promising discussions, in the event no comprehensive cases emerged, since data providers were unavailable or unwilling to progress discussions further within the time available. However, technical feasibility and outline business model approaches could potentially emerge in these areas, and similar considerations to those arising from the case studies below would apply.

The existing Linked Heritage data aggregation framework was proposed as a starting point for discussion; the results of this initial discussion are summarised in the table accompanying most of the case studies. Each table takes the form of the business model analysis in 4.4, with the colour of the table's frame indicating the acceptance or otherwise of the proposal in a red/amber/green traffic light scheme. Green indicates acceptance; red rejection and amber no final decision by the time of writing.

4.5.1 Heritage E-books (image library and associated e-book publisher)

Gwyn Headley
VisConPro Ltd t/a Heritage E-books
Murmur-y-Don
Harlech
LL46 2RA
Wales, UK

http://www.heritage.co.uk

info@heritage.co.uk

+44 1766 780263

This publisher is the smallest business considered here, having as yet only 40 titles on a well-defined cultural heritage topic (Follies of England). It is part of a larger business whose main focus is the sister company, an image library, some of whose content is used to produce and is showcased in the publishing wing's illustrated e-books. This publisher provided an excellent test case because:

- all rights in all types of content (the products themselves, preview images, metadata) related to the product are controlled by the publisher;
- the retail links can point directly to the retail platform owned by the publisher;
- all products have directly relevant content for Linked Heritage and Europeana;
- metadata could be sourced directly from the publisher on a manageable scale (only 40 records).

The publisher accepted the proposal to use the current Linked Heritage framework (as adapted by EDItEUR's Commercial Data Contributor Agreement) to publish metadata to Europeana, giving the reason that the data could be dedicated to the public domain without any problems and could only serve to promote the publisher's products.

Dimension	Sub-facet	Findings	Details / notes
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	For this publisher, a prototype ingestion is identical with a full ingestion under other full scale conditions, as they are willing to contribute their full title list. In this way it provides a realistic example, if not in quantity, at least in the quality of the records contributed
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Although this publisher disseminates their product data through other channels, including a major distributor and wholesaler who then passes the data on to various e-retailers, each has a retail page at the publisher's own website. As the publisher is also the data owner and has no reciprocal agreements controlling retail URLs with the distributor, they can decide which offering should be presented in their data
Service	Exposure	www.Europeana.eu	
	Coverage	Full dataset; 40 Records	Another 40 titles are planned but at present this is the full list. It is assumed that the new titles would also be included were they available within the lifetime of Linked Heritage

Dimension	Sub-facet	Findings	Details / notes
	Currency	One-off upload	No change to core descriptive data or basic availability likely since, the first set of 40 titles is complete and e-books are a secondary business to the image library. Further sets are planned, but this should not imply any change to the existing data.
Dataset	Product form(s)	E-books	The publisher's books are so far only available in electronic form, again, making this an ideal example in terms of relevance to the "digital cultural heritage" content remit of Europeana
	Classification(s)	All Records relevant	One very specific cultural heritage topic – *de facto* relevance.
Template	Record format(s)	ONIX 3.0	ONIX 3.0 Records created in-house by EDItEUR as no feed available; this took considerable time and effort as the publisher's data was in a SQL database exported as spreadsheets, and the available free, open source tools for producing ONIX were not set up for this specific task; hence a complex workaround, including direct editing (*e.g.* adding keywords, correcting logical inconsistencies in the data content) of the XML data, was used, requiring solid knowledge of ONIX and adaptation of existing tools
	Inclusion	All available ONIX fields , including	ISBN Title Author Publication date Product form Pagination Product language Classification codes Publisher Retail URL Keywords
	Granularity		Fully detailed ONIX records; some detail is lost in the translation to ESE
Other notes	Not all ONIX fields, even though successfully mapped to LIDO, could be mapped to ESE with the existing Linked Heritage MINT instance		

Although this publisher agreed to sign the Commercial Data Contribution Agreement, it should be noted that producing their ONIX records required significant work from EDItEUR, involving producing (using a template; see 13) and hand-editing the ONIX records for all 40 titles., Like many smaller publishers with short lists, this company does not provide an ONIX feed or subscribe to an ONIX-producing bibliographic database, their data was held in a proprietary format, the overall ONIX record creation involved multiple steps and would not scale.

This dataset produced attractive results in the MINT preview of the Europeana portal, complete with working cover images and links to the retail site. The ESE records are completely valid from Europeana's point of view; as bibliography *per se*, some details are not ideal, for example, the lack of a separate subtitle element (dcterms:alternative

was used instead purely for display purposes, to avoid concatenation into one dcterms:title, not because of a genuine semantic equivalence) and the current lack of an effective mapping of the product identifier in ESE (there is no clear distinction between identifiers for CHOs and DOs in ESE, and the LIDO mapping reflects this). However, the overall effect in the portal preview is adequate. For linked data, there would be some question as to the usefulness of the output given the lack of reliable identifiers.

4.5.2 Large image library

This large commercial image library made an excellent test case for Linked Heritage, since its collections are all highly relevant to Europeana's content development policies. It holds its own collection of unusual historical images, represents over 350 specialist collections with some sort of historical, cultural or heritage focus, such as social history, modern art and design, and content from large historical and institutional archives and several museums and heritage organisations.

However, though they showed a clear interest in the project and involvement with Europeana in general, they did not wish to contribute any data under the current conditions; the primary reasons given were:

- that the CC0 requirement in the Europeana DEA would remove control over their database of image identifiers and keywords, which constitute a valuable research asset in themselves;
- that CC0 would remove the possibility of updating the commercial aspects of the data, primarily, their availabilty from this source (see 4.4.8 for a full discussion of these issues);
- other areas of concern about the technical aspects of the Europeana portal, centred on protection of, and attachment of metadata to, image thumbnails in Europeana (these are detailed more full in section 4.2.4).

Although of course no concrete partnership was proposed in this case, the "business model" table below summarises some aspects of the image library's capabilities to highlight what *could* be available under *different* licensing and technical conditions.

Dimension	Sub-facet	Findings	IMAGE LIBRARY 1
Partnership		Proposed agreement via EDItEUR (Linked Heritage) not acceptable in principle.	
Offering	Retailer(s)	Library's own platform; derivative products through prints company.	
Service	Exposure	www.Europeana.eu	
	Coverage	Approximately 300,000 (current number of existing digital scans; materials not yet scanned make up considerably larger numbers).	
	Currency	n/a – no ingestion possible in principle, but technical capacity for frequent updates.	
Dataset	Product form(s)	Images, prints and various "merchandise" *e.g.* cups, shirts, computer accessories based on photo prints	
	Classification(s)	Proprietary keywords exist but all content is relevant to Linked Heritage.	

Dimension	Sub-facet	Findings	IMAGE LIBRARY 1
Template	Record format(s)	IPTC	
	Inclusion	Title, keywords, retail URL.	
	Granularity	n/a	
Other notes	The “title” of many products may be in reality a short description of the content of the image, rather than a title proper; this is in keeping with the heritage sector data format LIDO, although it does not translate well into Europeana’s ESE format.		

This image library has several standard contracts, which, in addtion to technical protections detailed in section 4.2.4, include stipulations which are not present in the Europeana DEA:

- Territory / application context (*e.g.* Europeana.eu with world-wide access; for linked open data under CC0, this can by definition not be specified);
- Arrangements for promotion;
- Compensation (financial);
- Exclusivity (impossible under CC0);
- Non-transferability (impossible under CC0);
- Warranties and transfer of warranties (impossible under CC0);
- Rights to make versions of preview images;
- Deletion of data on expiry/termination of contract (impossible under CC0);
- Guaranteeing removal of incorrect data (especially regarding availability);
- Indemnity from consequences of non-removal or correction.

One key component of this library’s business model is the ability license digital images created from, and significantly improving on, analogue (printed or negative) originals, many of which originals (but not all) will be in the public domain. Products derived from these original artefacts represent added value due to the skilful processes used to produce images which are clearer, have richer colours, corrections to artistic or printing defects, or due to wear and tear of the physical medium; in short, they represent new images which are copyright protected.

Licensing of these new images provides revenue not only in order to recoup the original production costs but to also ensure the stability and growth of the image library, and hence the preservation of its collections of original artefacts. Many of these highly specialised collections, though culturally valuable, may not qualify for direct public funding for preservation and exhibition; thus this business model could be seen as having an aspect of heritage preservation and promotion to the benefit of both the public and commercial sectors.

Europeana[101], Wikimedia Foundation[102] and others[103] have made strong statements on the status of digital versions of public domain objects and images, which focus mainly on unenhanced digital photographs. Europeana's Public Domain Charter (footnote 101), for example, says:

> "Works that are in the Public Domain in analogue form must remain freely available in digital form…"

However, it is not immediately clear what are the limits within which a work in digital form is substantially the same as in analogue form, which may lead to concern from image libraries that their product may be seen in this light as "simply a faithful reproduction" (see footnote 102) rather than a work in its own right.

There may be a need to more emphatically and clearly reassure commercial players, like this image library and others like it, that these statements are not a condition of supplying data to Europeana, and will not be construed to suggest that rights in the full product content will not need to be relinquished as well. Europeana's Charter does contain a helpful clarification:

> "This Charter is a policy statement, not a contract. It does not bind Europeana's content providers to any position."

It may also help in future to clarify that the Charter is intended for the use of primarily publically-funded "memory organisations", and recognise the complementary contribution of cultural content curators that operate with different business models and different policy needs.

[101] http://www.europeana.eu/portal/rights/public-domain-charter.html
[102] http://commons.wikimedia.org/wiki/Commons:Policy_on_photographs_of_old_pictures
[103] For example: http://www.publicdomainmanifesto.org/manifesto#General

4.5.3 Large academic library's publishing arm

We contacted the sales and marketing coordinator of a publishing section within a major academic library. The library combines working university library services for a top world university based in the UK with extensive special collections that contain manuscripts and rare books from Western Europe and worldwide, many of which have been digitised, and a significant number published as digital or print fascimilies, or made the subject of authoritative, scholarly treatments or popular introductions. These published books are produced and sold through the publishing company wholly owned by the university.

This case is a clear example of the "content selection by provider" model, in that products from this publisher are based in some way on the collections and expertise of a hertiage institution, so they are *de facto* relevant to Linked Heritage, according to Europeana's content development policy. Europeana should note that publishers like these represent the purest expression of commercially publishing, enriched cultural content available through online channels, as well as supplying some of the richest and most culturally relevant metadata possible, thanks to their use of specialised data service providers.

With explicit reference to the heritage and educational mission of their parent organisation, this publisher responded with a tentative "yes" to the Linked Heritage proposal, pending resolution of one technical and organisational aspect, namely, the persistence of retail site URLs and the ability to update these in Europeana's portal and linked data publications beyond the lifetime of the Linked Heritage project.

The ONIX 2.1 files provided for this publisher were supplied by the book data service described in case 4.5.8; since that service acts only as a data platform and distributor, with the publisher retaining the rights in the data itself, all business decisions were discussed directly with the publisher, and all technical points could be addressed through the data service company.

For small to medium-sized, highly focussed cultural heritage publishers like this one, working directly with the publisher in combination with their technical supplier appears optimal; but note that substantial time and specialised work were invested to coordinate discussions, solve technical queries and explain the terms of the agreement. This is not a zero-effort option, merely the one with the most direct routes to solving any problems, and the highest potential return in terms of data relevance and richness.

Dimension	Sub-facet	Findings	Details / notes
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	Complete acceptance in principle; DEA seen as compatible with educational and heritage mission of the publisher; but see below for concerns with currency.

Dimension	Sub-facet	Findings	Details / notes
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Although the data supplier was easily able to include retail URLs for each of the publisher's products, resolving to pages at the publisher's own retail site, the parent academic library organisation also has a separate webshop which sells the publisher's products in the context of visits to the library's collections (online or in person). It might be reasonable to attempt to include both, but note that this would entail further work to ascertain which products appear in each retail context, attach the correct URL to each, and ultimately, the publisher themselves noted, update those links in the Publication Subset.
Service	Exposure	www.Europeana.eu	
	Coverage	Full dataset	This figure represents the current list of in-print titles; a further set of records for the publisher's backlist could be created either by the data service provider, or by EDItEUR from spreadsheets (as in case 4.5.1) but this decision was not completed by the time of publication.
	Currency	One-off upload	As already noted, concern over updating the data has delayed expression of firm support for upload .
Dataset	Product form(s)	Print, CD-ROM	Many digital manuscript facsimilies with enhanced features are only available on CD-ROM.
	Classification(s)	All Records relevant (BIC codes available)	Publisher attached to major university's library and archive collections.
Template	Record format(s)	ONIX 2.1	ONIX 2.1 feed created by publisher through dedicated book data service.
	Inclusion	All available ONIX fields, including:	ISBN Title Author Publication date Product form Pagination Product language Classification Publisher Retail URL Descriptions
Other notes	Not all ONIX fields, even though successfully mapped to LIDO, could be mapped to ESE with the existing Linked Heritage MINT instance.		

4.5.4 National library's publishing arm

We contacted the sales, publishing and brand licensing manager of the wholly-owned publishing section of a major national library. The library publisher in this case study has many similarities with the major academic library publisher in case 4.5.3:

- The national library also publishes content that is *ipsofacto* relevant to Europeana, consisting of facsimily editions of culturally important works, new editions of published books, selections from various national audio and visiual archives, and original performances of texts, scholarly appreciation and introductory guides; being a national repository for cultural artefacts, it also publishes a large number of exhibition catalogues and guides to material and social book and publishing culture as such – but in contrast to the academic library, many of the national library publisher's books are co-published, a common practice among heritage institutions;
- The national library's publishing operation also uses a third-party data service to produce its ONIX records, meaning that technical aspects can be addressed separately from legal-commercial questions;
- There is a webshop attached to the main library website which also sells a variety of non-book products including art and photography prints and gifts.

In this case, the publisher responded with a firm "yes" to the proposal of contributing data, pending availability of an ONIX 2.1 or 3.0 feed from their outsourced data platform supplier.

Dimension	Sub-facet	Findings	Details / notes	NATIONAL LIBRARY
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	Complete acceptance in principle; DEA seen as compatible with educational and heritage mission of the library, which already has some partnerships with Europeana	
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Retail links can easily be programmatically created from a base URL plus product ID, the latter being present in the ONIX data	
Service	Exposure	www.Europeana.eu		
	Coverage	100 current titles (approx.) plus around 100 backlist	This figure represents the current list of in-print titles with retail pages in the national library's webshop	
	Currency	One-off upload		

Dimension	Sub-facet	Findings	Details / notes	NATIONAL LIBRARY
Dataset	Product form(s)	Print books		
	Classification(s)	Not known.	All products relevant for Linked Heritage	
Template	Record format(s)	ONIX 2.1 or 3.0 (TBC)		
	Inclusion	All available ONIX fields, including *e.g.*	ISBN Title Author Publication date Product form Pagination Product language Classification Publisher Retail URL Cover image URL	
Other notes	Cover image URLs can, like retail URLs, be generated from the ISBNs of products in the webshop for this publisher			

4.5.5 Large academic e-book retailer

We contacted the head of digital product development at a major European publisher, which also runs a distinct e-book retail platform. An example of a mixed business model, where the e-book retail platform is owned by a company that acts as a publisher, but also a distributor and retailer of its own and other publishers' titles. The e-book platform already makes freely available basic structured data about its full title list as an Excel spreadsheet with column headings as below:

Publisher	Local ID (ordering number)	Title	Author	Publication year	ISBN	Dewey Decimal Classification number	Dewey Decimal subject heading	URL to retail web page

This means that in principle the data could be downloaded, programmatically converted to ONIX (using the recipe in section 13), LIDO or directly to ESE / EDM, and uploaded to Europeana without much extra effort than in cases 4.5.1 and 4.5.6; the data is already *technically* "open", but, especially as no terms of use are mentioned on the website or within the file itself, of course permission should be sought from the retail platform provider who owns the data and controls the right to redistribute it in other forms (2.1.3), and should be considered to grant an implicit licence as to how it should be presented, even as they offer it on the open web (see 3.1).

One mandatory field for Europeana (see 4.2.1) is the language of the text, not given in the open spreadsheet; this would have to be obtained separetely from the retail platform, or perhaps could be added to the spreadsheet at some future time.

This case study is classified as a qualified "yes" to Linked Heritage's proposal since the contribution of data is agreed in principle, but pending production of the ONIX data feed.

Dimension	Sub-facet	Findings	Details / notes E-BOOK PLATFORM
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	Complete acceptance in principle; Europeana DEA seen as compatible with eductional mission of retail platform with primarily academic customers
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Retail links can easily be programmatically created from a base URL plus product ID, the latter being present in the ONIX data, or copied directly from the URL in the data feed (the latter may be better, since some products do not have a standard ISBN)
Service	Exposure	www.Europeana.eu	

Dimension	Sub-facet	Findings	Details / notes E-BOOK PLATFORM
	Coverage	2468 records	Of roughly 10,525 titles, approximately 2,500, or just over 20% were deemed relevant, based on a basic, approximate selection (found in Appendix section) of Dewey Decimal Classification headings chosen by Linked Heritage WP4 from those present in the open catalogue listing
	Currency	One-off upload	
Dataset	Product form(s)	E-books	
	Classification(s)	Dewey Decimal Classification	Codes and subject headings available for almost all titles (only 18 lacked subject headings and/or codes)
Template	Record format(s)	ONIX 3.0 (TBC)	
	Inclusion	All available ONIX fields, including *e.g.*	ISBN Title Author Publication date Product form Pagination Product language Classification Publisher Retail URL
Other notes	Cover image URLs cannot be included because of the need to request permission from the individual publishers		

The terms and conditions for publishers contributing their title data (and product content) to this platform include:

- Scope:
  - Online retail;
- Allowed and forbidden uses (by retail platform service):
  - Communication of data to other aggregators with permission from publishers;
  - Marketing and publicity activites;
- Duties (of publishers contributing data):
  - At least annual updates;
  - Cooperation to create bibliographic data (“keywords”).

The agreement does not specify uses of the bibliographic data itself, but these could be considered part of the marketing and promotion activities which are mentioned, and include, for example, full-text search and first page preview display, which have far wider scope than most bibliographic data.

4.5.6 Large national academic e-book publishers

We contacted the data service managers and rights Two of the largest university presses in the United Kingdom are known to have openly accessible data services, though both require free registration. Both produce, among other output formats, ONIX 2.1 feeds. Press A's service has the most open terms of use and is presented here first.

According to their website, this is "a free resource for downloading product information for titles published and distributed by" Press A; "Who can use it? Anyone with access to the internet - whether bookseller, librarian or distributor. The service is free."

This appears to mean that Europeana (or the Linked Heritage coalition) could technically, and legally, harvest book records for upload to the portal and redistribution without needing specific permissions, although without any guarantee for future access since Press A "reserves the right to decline or curtail access… at its own discretion."

Dimension	Sub-facet	Findings	Details / notes	PRESS A
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	Because the product data are offered with essentially no explicit terms and conditions, beyond registering with the Web service to obtain access, it is especially difficult to characterise this case as a "partnership". Confirmation that Linked Heritage's envisaged use (ingestion, LIDO mapping and publication of a subset to Europeana) would be acceptable was obtained via email, and EDItEUR's prototype agreement proposed.	
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Retail links can easily be programmatically created from a base URL plus product ID, the latter being present in the ONIX data.	
Service	Exposure	www.Europeana.eu		
	Coverage	Variable	Potentially hundreds of thousands of records are available from this publisher; several attempts to create relevant subsets of their in-print (or, for e-books, in-commerce) title list resulted in almost unmanageable filesizes; see below for more details.	
	Currency	One-off upload	The number of records retrieved was extremely large, and a dummy "provenance" record is appended to the start of every download from this service. Checking the real records and removing the dummies would be highly resource intensive and so a one-off prototype upload only was performed. In principle regular FTP updates of only the new/changed records only (still probably a large number) would be technically feasible with the existing infrastructure.	

Dimension	Sub-facet	Findings	Details / notes	PRESS A
Dataset	Product form(s)	Print and e-books	Both print and e-book formats are available in one dataset; or records can be filtered by product form to make datasets more manageable.	
	Classification(s)	Proprietary retail classification (based on BIC)	The proprietary subject headings available for dataset filtering in this service are more general than the BIC classification, and do not allow general geographic qualifiers. Therefore many headings that could be relevant when qualified with European focus were omitted to preserve relevance. Even with this minimal approach, as noted above, the total number of records was high.	
Template	Record format(s)	ONIX 2.1	Rich ONIX 2.1 with a large quantity of author information, subject headings, series and component titles and descriptive texts / summaries.	
	Inclusion	All available ONIX fields, including *e.g.*	ISBN | Title | Author | Publication date | Contributor descriptions | Reviews | Product form | Pagination | Product language | Classification | Publisher | Retail URL | Cover image URL	
	Granularity	For the academic titles in this publisher's list, many records have details of component parts, such as chapter headings and volume titles, which are highly valued by scholarly readers; this detail can be rendered accurately in LIDO but loses its indexing value through translation to ESE, where it is only useful for basic free-text search. It remains to be seen how well such "internal" description fields, as well as relations such as review events and contributor actor descriptions[104] , will map to EDM, which is currently under review with the aim of enhancing its bibliographic capacity[105].		
Other notes	Not all ONIX fields, even though successfully mapped to LIDO, could be mapped to ESE with the existing Linked Heritage MINT instance.			

The publilsher was contacted directly to confirm this position; the reply confirmed the in-principle acceptance of Linked Heritage's proposal:

> "there are no terms & conditions regarding sharing [Press A's data service] output with third parties. However sample content is copyrighted so customers can only use this for promotional purposes & cannot quote directly from it without [first clearing the rights]…"

[104] See detailed coverage of these mapping problems in D4.2, available at: http://www.linkedheritage.org/getFile.php?id=394
[105] See EDM-FRBRoo enhancement task force activity at: http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+EDM+FRBRoo

"Everything you have described below [*i.e.* the Linked Heritage proposal in outline] is perfectly acceptable to us."

However, this publisher was not able to finally confirm or refuse acceptance of the specific licence terms of the adapted contributor agreement (Appendix section 12) at the time of writing; this appears to reflect the fact that Linked Heritage's model of reuse was not envisaged by this publisher (in common with many others) when setting up the data service – it is primarily intended for familiar categories of reuser, "bookseller, librarian or distributor". For this reason Linked Heritage WP4 has not uploaded any prototype datasets from Press A to date.

The second university press offers a data service comparable to that of Press A, except for somewhat more detailed and varied output format options, and explicit terms and conditions on their website, for some types of content such as cover images and book previews. As in the case above, the university press was approached directly to discuss the outline proposal and the specific terms of Linked Heritage's modified contributor agreement. The publisher responded:

"…the first question that needs to be asked is a business one rather than a legal one, of whether or not the Press wishes to enter into any such agreement…"

A later reply was fundamentally positive:

"the academic team have provisionally confirmed that this would be a project that they would be interested in."

As with the previously described case, at the time of writing there appears to be no obstacle to sending Europeana textual bibliographic data provided at no charge by Press B, though explicit terms of use prohibit redistribution of their book cover images. The image URLs would have to be left out of the Publication Subset of data.

This case confirms the impression received from the other university press data service; that this type of reuse is unforeseen by these publishers, but in principle, and most likely in practice, acceptable. It appears unlikely that such reuse would be refused by the presses should it be directly proposed, but they are unlikely to see the need, or be able, to enter into formal agreements for supply. Hence it might also be noted that no guarantees, either for sustained future data supply, or for availability of the products referred to in the data, might be obtained by heritage organisations wishing to reuse these data sources "as is".

Dimension	Sub-facet	Findings	Details / notes	PRESS B
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)	As for Press A, no "partnership" in the sense of a formal agreement, appears necessary or even truly possible, since the service is offered openly, terms of reuse for images notwithstanding.	
Offering	Retailer(s)	One: direct retail link at publisher's own Web site	Retail links can easily be programmatically created from a base URL plus product ID, the latter being present in the ONIX data.	
Service	Exposure	www.Europeana.eu		

Dimension	Sub-facet	Findings	Details / notes	PRESS B
	Coverage		As for the other university press, test downloads produced extremely large numbers of records even with strict filtering by subject and publication date; although not a heritage institution *per se* a very high proportion of Press B's content is relevant	
	Currency	One-off upload	As for Press A, a one-off upload, probably in several parts for convenience , would be practical for obtaining a large number (hundreds of thousands) of relevant, recent publications	
Dataset	Product form(s)	Print and e-books		
	Classification(s)	Proprietary retail classification (based on BIC)	The proprietary subject headings available in this service are more general than the BIC classification, and do not allow general geographic qualifiers. Therefore many headings that could be relevant when qualified with European focus were omitted to preserve relevance. Even with this minimal approach, as noted above, the total number of records was high.	
Template	Record format(s)	ONIX 2.1	Rich ONIX 2.1 with a large quantity of author information, subject headings, series and component titles and descriptive texts / summaries.	
	Inclusion	All available ONIX fields, including *e.g.* ISBN Title Author Publication date Contributor descriptions Reviews Product form Pagination Product language Classification Publisher Retail URL		

As with Press A, this university press data service's website explicitly states that its service is universally open and free of charge:

> "There is absolutely no charge to use [the service]… Anyone with access to the internet [may use it]. There is a simple registration process and you are identified by email name."

The terms of use for cover image files are displayed within the file specification interface and must be accepted in order to proceed with download:

> "I understand that the material I download from [Press B] is in copyright and is made available for the purpose of promoting/publicising the title(s) to which it relates, and/or for personal use, and that it may not be sold or used for other purposes."

Any redistribution of the cover image files through Europeana would have to proceed on a basis whereby Press B would understand display of the images in the Europeana.eu portal as "promoting/publicising the title(s)" to which they relate.

4.5.7 Large international academic e-book publisher

The Linked Heritaga proposal below was discussed with the relevant senior executive at a large academic publisher, with reference to one of its humanities imprints, in the context of linking product records to Web pages within the publisher's own retail platform.

With the publisher's help, a subset of the entire print and e-book offering (just less than 3% of products) was specified as directly relevant for inclusion in Europeana, using the subject classification method detailed in section 0. The titles and product formats (print or e-book) were reviewed and confirmed as highly relevant to cultural heritage.

Even this minimal proposal was declined. The stated reasons for declining the agreement hinged on the CC0 rights and licence waiver for redistribution of data; for this publisher, any type of Creative Commons licence would be unacceptable, and especially the CC0 code. This was because of

a) the terms of these licences cutting directly across existing contracts with clients of the publisher guaranteeing a revenue stream from the supply of this data (normally the full dataset rather than the subset proposed by EDItEUR) and;
b) the (theoretical) possibility of enabling replication of existing services (even were the template restricted to just ISBNs and retail Web page URLs) with no legal recourse.

This publisher remained open to further discussion should a new licensing framework become available.

Dimension	Sub-facet	Findings	PUBLISHER 1
Partnership		Proposed agreement via EDItEUR (Linked Heritage)	
Offering	Retailer(s)	One; publisher's own retail platform	
Service	Exposure	www.Europeana.eu	
	Coverage	Subset: 2.6% of all records from humanities imprint (counted by product, not by title, though the percentage would be on the same order of magnitude for titles); this corresponds to roughly 2.5 thousand relevant titles	
	Currency	n/a (hypothetical prototype ingestion only)	
Dataset	Product form(s)	Print and e-books	
	Classification(s)	BISAC subset (see section 11)	
Template	Record format(s)	ONIX 2.1	

Dimension	Sub-facet	Findings	PUBLISHER 1
	Inclusion	ISBN, Retail URL (other core fields technically available but publisher unable to discuss specifics under CC0 redistribution terms)	
	Granularity	n/a	
Other notes	No cover images included as redistribution rights not held in any covers containing photographs or other images (many titles have a generic background design but this is not noted in the bibliographic data).		

It should be noted that large publishers will be an obvious starting point for large numbers of relevant product records, and that as in this case, they are likely to have existing commercial relationships with libraries, distributors and retailers, which involve metadata as well as product content, often because, quite simply, libraries will not have the available resources to produce new catalogue data for large numbers of digital products in-house. Therefore, the library buys in trustworthy data from publishers or other commercial supply chain partners. This follows established patterns for library-publisher relationships[106], which will have to be the starting point for any successful negotiation in this area.

[106] For a technical overview of existing commercial data supplier cooperation with libraries, see Bell, 2013: http://leo.cilea.it/index.php/jlis/article/view/5487

4.5.8 Medium sized book data service providers

This data service provides a full technical implementation of ONIX 2.1 for over 80 clients, who are publishers of a variety of sizes and degrees of specialism. While the technical platform is provided by the data service, full rights in the data and control over its distribution are held by the publishers themselves. This is an example of the separation of service provision from data ownership discussed in section 4.3.5.

Many of the service's clients are university presses, or specialist publishers on specific academic subjects. Others are small and medium general publishers, often with one or several specialist series or imprints. Out of these 80+ publishers, around 30, or 40%, had partly or fully relevant title lists. Unfortunately it is not possible to filter these lists by subject category, so only publishers with wholly relevant title lists were approached (being the data owners, the business decision as to whether to contribute data is in their hands).

Although the academic library publisher in case 4.5.3 was one of only two such publishers directly contacted by Linked Heritage, it must also be noted that similar publishers, such as university presses, publishing wings of public heritage bodies, or independent publishers of academic studies on historical and cultural topics make up a large proportion of this particular data service's clients, making the data service an excellent contact to obtain permission from other publishers to receive product data. The more publishers such a data service has as clients the better, since all of these specialist publishers are likely to offer only a numerically limited number of product records each, but having a central contact point for them all in the data service means that agreements with all of them can be proposed more efficiently and the combined data contribution could be reasonably significant.

One of the specialist small presses was contacted incidentally by Linked Heritage and expressed clear acceptance in principle of the proposal to contribute data; this reinforces the impression that only the smaller commercial publishers (like the one in case 4.5.1) are likely to have no commercial barriers to contributing their often highly relevant but generally extremely small datasets. In the final analysis, having to reach separate agreements with each of a large number of data owners will limit the scalability of dealing with purely technical aggregators in the Europeana context.

4.5.9 Large national book data service providers

The book data services described here are all managed and maintained by the national publishing and book trade associations in their respective countries, rather than fully independent companies. They each have unique profiles (see section 5 for a detailed discussion) in terms of their own technical functions and their legal-commercial relationships with other services offered by their trade body and client base.

National Service 1

The National Service 1 case describes two separate data ingestions from the same service. The first ingestion chronologically, mapping direct from ONIX 2.1 (the native data format of the service) to ESE, was negotiated separately with Europeana by the national service provider before and during the lifetime of Linked Heritage, while the second ingestion is planned as a Linked Heritage WP4 prototype ingestion, mapping via LIDO. This explains some small differences in the details of the ingestions noted in the table below.

Importantly, as yet no data has been contributed to Europeana, either directly or via Linked Heritage. This may be partly due to technical difficulties, but also to the difficulty of reconciling the standard agreement, described below, to Europeana's DEA.

Publishers may usually contribute data to the service free of charge (although they may be members of the trade association and thus support it indirectly through their membership fees). In normal use, National Service 1 applies a contract with strict terms of use for the data *recipient*, specifying, among other terms:

- Scope:
  - Online retail;
- Allowed and forbidden uses:
  - Search by, display to, and ordering by customers;
  - NO redistribution or provision of public access (beyond single title display or recommendation lists) to the whole database or substantial portions of it;
  - NO replication of service and especially NO creation of competing database products incorporating the database.
- Duties:
  - Recognition of National Service 1's copyrights and database rights in the database;
  - Requesting written permission for any use beyond those explicitly allowed in the contract;
  - Deletion of the user's copy of the database upon termination of the contract.

Many of the terms are also enforceable by termination of the contract and financial penalties specified in the contract. The annual cost of *receiving* the service's data is around €10,000. This reflects the value, primarily for retailers, of receiving an up-to-date and therefore comprehensive data feed – a data *service*, rather than only a snapshot.

Dimension	Sub-facet	Findings	Details / notes	NATIONAL SERVICE 1
Partnership		Direct agreement with Europeana (DEA)	For the Linked Heritage prototype ingestion, WP4's modified contributor agreement was used.	
Offering	Retailer(s)	One; multi-publisher platform owned by parent company of data service		
Service	Exposure	www.Europeana.eu		
	Coverage	Full dataset; 180,000 Records	Applies to both ingestions in principle; for LInked Heritage ingestion only relevant titles selected.	
	Currency	Monthly updates	Linked Heritage ingestion is a one-off prototype upload.	
Dataset	Product form(s)	Print and e-books	Again, in principle all records; Linked Heritage requested only e-books, filtered by ONIX product form (this is not possible using the combined classification system; see below).	
	Classification(s)	Warengruppen 2.0	Classification combines product form and subject heading in each code; selection was made using only subject codes (subjects are identical for any product form range so in theory this would apply to print books, audio-visual materials *etc.*). The list of relevant WG codes is found in Appendix section 11.4.	
Template	Record format(s)	ONIX 2.1		
	Inclusion	Ten specific fields:	ISBN Title Author Publication date Product form Pagination Product language Classification Publisher Retail URL	
	Granularity			
Other notes	No cover images included since redistribution rights not owned by the service			

The ingestion defined above is offered to Europeana under the standard DEA terms and thus fulfils almost none of the requirements of the usual licence agreement. There is a formal agreement to provide the ten data fields above, and an (informal) expectation to offer ongoing monthly updates for the full database. The ten fields listed could certainly fulfil the minimum requirements of the ESE/EDM specifications; the DEA makes no demands in terms of updates or quantity of records.

This is therefore a limited cooperation with no formal commitment to provide any specific set of records. The maximum possible comittment for a commercial data provider is apparent in this case; another crucial factor is the presence of a *single central retail platform for all products* described in the database, so that there is no question of commercial sensitivity relating to choice of retail link for any given record. Without this it is doubtful if any agreement could be made.

National Service 2

National Service 2 has a similar functional profile to that of National Service 1 except that it does not relate to any central retail platform. In fact, no retailer links are present in the ONIX data provided by this service. The stated reason for this is that:

- No retailer links are stored in the database anyway (only distributor information is provided);
- Adding retailer links to the data contributed in this project is not desirable because National Service 1 does not wish "to appear on Europeana as "data provider" [representing] a single [retail] provider";
    - This is partly because National Service 1's "bibliographic description and cataloguing could be different from that on the retailer's site";
    - Linked to this, there is no formal technical mechanism for directly harvesting publishers' metadata, and hence no formal agreement that data contributors to National Service 2 would sign.

Another difference is that cover image links *can* be included in principle. This is reasonable considering that no formal agreement with providers exists concerning these image files for National Service 2, and that it is possible to prohibit redistribution of thumbnail versions (and links to them) from Europeana's server.

The case study has been counted as an "acceptance" of the existing terms, even though there is clearly less commercial consideration in return for the metadata received by Europeana should this ingestion go ahead, since the books cannot be directly purchased from the Europeana search result page, unless in future an ID resolution service is implemented (for example, ISBN-A, as in section 5.1.3, or some other meta-resolver as in section 8.1.3).

Dimension	Sub-facet	Findings	Details / notes	NATIONAL SERVICE 2
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)		
Offering	Retailer(s)	None	Product records only, showing that relevant products are available and uniquely identifying them.	
Service	Exposure	www.Europeana.eu		
	Coverage	3600 e-books; 87000 print books	These are rounded figures; note that the 3600 e-book product records represent approximately 2900 different titles since, following industry best practice, each different e-book file format, e-reader platform or DRM type offered for a given "work" represents a new *product*, even if the *content* is the same.	

Dimension	Sub-facet	Findings	Details / notes	NATIONAL SERVICE 2
	Currency	One-off upload	Updates to product records beyond the Linked Heritage project lifetime are not feasible due to the expense of maintenance.	
Dataset	Product form(s)	Print books and e-books (separately)		
	Classification(s)	CCE 2-2.0	Content selection as per BIC with the addition of CCE-specific codes (see section 13.3 below for the full listing)	
Template	Record format(s)	ONIX 2.1		
	Inclusion	Five core fields agreed for Publication Subset, with an extra field for e-books:	ISBN Title Author Publication date Publisher Cover image URL Product form detail (e-books only) Publisher description (possibly)	
	Granularity			
Other notes	Subject classifications (codes and subject headings) are routinely added to ONIX files by this data service, but these are not included in the Publication Subset, in order to protect a significant value-added enrichment provided by this service, and as they are not seen by the service as appropriate for transmission into a library-like environment where typically other, very different classification schemes are used.			

The normal terms of use for National Service 2 are also strict, although without financial penalties as in the previous case:

- Scope:
  - Online book retail with the user's own clients;
- Allowed and forbidden uses:
  - Keeping one copy of the database, exclusively on the server identified in the contract;
  - Granting access to the data only on Web sites and terminals linked to the identified server;
  - Search and display to clients on Web sites identified in the contract;
  - Internal use for ecommerce-related activities;

- NO redistribution, sale, licensing or provision of access to the database;
- NO creation of any copies beyond one back-up copy.

- Duties:
  - Recognition of National Service 2's copyrights and database rights in the database;
  - Requesting written permission for any use beyond those explicitly allowed in the contract;
  - Display of a specified copyright (and all other rights) and attribution statement next to any display of database contents and cover images;
  - Deletion of the user's copy of the database upon termination of the contract.
  - Specifically for images supplied via the database:
    - NO modification of any kind (*e.g.* change of size, quality, file format…);
    - Display only in the supplied format;
    - NO use apart from with bibliographic descriptions;
    - Removal upon request of copyright-infringing images or those retracted by the copyright owner;
    - Deletion of the user's copy of the images upon termination of the contract;

This is an example of a books in print service for which publishers do not pay for their products to be listed; as such it would be a perfect test case (though not necessarily the only one) for the business models and incentives sketched out in section 5.

National Service 3

The functional profile of National Service 3 is quite different from the others considered here, because it combines aspects of a books in print service with aspects of a publisher data service; each data contributor retains separate control over their data contribution and access to the overall data aggregation is (currently) free of charge, although a stringent user licence still applies, and paid agreements exist with book data services that, in turn, sell their ONIX data to libraries (*i.e.* commercial reusers of the data – see also the commercial rationale behind this expressed in 4.4.5).

National Service 3 provides in total over 471000 records, of which approximately 316000 (*i.e.* most records, around 67%) have BIC subject headings and thus could be filtered for relevance to Europeana. This also provided a good test case for the subject classification method of content filtering since this test yielded the following results:

- Directly relevant subject headings: 4756 records;
- Headings deemed by WP4 to be relevant with EU qualifiers:
  - Combined with qualifier, 1219 records;
  - Without qualifier, 62,401 records.

The comparison of qualifier combinations was suggested by the data service providers themselves, since the accepted industry practice was known to assume that subject headings referred to this EU state unless another country was indicated. That assumption was borne out by the data. Further data sampling and testing along these lines could be justified in future projects, especially with regard to the new international subject listing, Thema[107], which will possibly facilitate discovery and purchase of books across language barriers.

At the time of writing, the Linked Heritage proposal could not be accepted in principle, since the data owners, the data contributors themselves, would have to give consent to this on a case-by-case basis. This does not rule out more complex agreements in future projects.

Dimension	Sub-facet	Findings	Details / notes	NATIONAL SERVICE 3
Partnership		Contribution to Europeana (DEA) via EDItEUR (Linked Heritage)		
Offering	Retailer(s)	None.		
Service	Exposure	www.Europeana.eu		

[107] http://www.panthema.org/

Dimension	Sub-facet	Findings	Details / notes	NATIONAL SERVICE 3
	Coverage			
	Currency	One-off upload		
Dataset	Product form(s)	Print and e-books		
	Classification(s)	BIC		
Template	Record format(s)	ONIX 3.0		
	Inclusion	All available ONIX fields, including *e.g.*	ISBN Title Author Publication date Contributor descriptions Reviews Product form Pagination Product language Classification	
	Granularity			
Other notes	Only approximately 25% of cover images have URLs referenced in the ONIX data.			

The licence terms for National Service 3 are significantly more concise than those for the other data services but in effect equally restrictive in terms of reuse:

- Scope:
  - Online book retail with the user's own clients.
- Allowed and forbidden uses:
  - Incorporation of the database into the user's own database;
  - Provision of public access to user's database;
  - NO redistribution, sale, licensing or third-party provision of public access to the database.
- Duties:
  - Recognition of National Service 3's rights in the database.

It is worth noting that this service does not charge for access to its data but does not rule out charging for the service in future. Therefore although commercial reuse is ruled out for most users, this is to protect the integrity of the database and supporting services on behalf of its users, apart from the need to preserve revenue streams from use charges. It is thus a mid-way case in terms of funding model too, and as such may be open to a range of partnership proposals in future projects, should the CC0 redistribution requirement change.

National Service 4

Another national book data service offers a simple-to-use REST API[108] which accepts searches with parameters including, *e.g.*, date of last change to the data, availability status, (ranges of) subject codes[109], (ranges of) publication dates, language codes[110], publisher, imprint[111], or (list of) ISBNs.

The service outputs an extremely simple record format that can be almost trivially mapped to ESE/EDM, for example as:

National Service 4 output format	ESE
<product>	<record>
<productid>640116</productid>	<dc:identifier>640116</dc:identifier>
<isbn>9789461400208</isbn>	<dc:identifier>urn:isbn:9789461400208</dc:identifier>
<titel>Jan Jansen architecten</titel>	<dc:titel>Jan Jansen architecten</dc:titel>
<auteur>Kloos Maarten, Kleijn Koen</auteur>	<dc:creator>Kloos Maarten, Kleijn Koen</dc:creator>
<medewerker>Bekkers Gaston</medewerker>	<dc:contributor>Bekkers Gaston<dc:contributor>
<uitgever>Architectura & Natura</uitgever>	<dc:publisher>Architectura & Natura</dc:publisher>
<imprint>Imprint</imprint>	As noted above in footnote 111, “Imprint”, in publishing, is a brand name used by a publisher, and may be transferrable, and so is not mapped to dc:publisher here.
<pubdate>20111123</pubdate>	<dcterms:issued>20111123</dcterms:issued>
<prijs>15,00</prijs>*	A product’s price is commercial, rather than bibliographic, information, and also one of the most changeable elements of product data; therefore it is strongly recommended not to provide it to Europeana at the current level of update frequency (see discussion of updates in section 4.4.8).

[108] See *e.g.* https://www.ibm.com/developerworks/webservices/library/ws-restful/ for background on REST.
[109] In this case, NUR; see http://www.boek.nl/nur
[110] From ONIX code list 74; see http://www.editeur.org/ONIX/book/codelists/current.html
[111] See *e.g.* http://www.sherpa.ac.uk/romeo/publishertypes.php#imprint or http://slb-ltsu.hull.ac.uk/awe/index.php?title=Imprint

National Service 4 output format	ESE
<cover>http://link.be/naar/cover.jpg</cover>*	<europeana:isShownBy>http://link.be/naar/cover.jpg</europeana:isShownBy>
<nur>648</nur>	<dc:subject>Bouwkunst, architectuur</dc:subject>
<notificationtype>01</notificationtype>	Records should be filtered on this field to allow only published books to be represented (code 03 from ONIX code list 1).
<productavailability>20</productavailability>	Products' availability should be used as a filter to select only books currently available to retail.
</product>	</record>

Note that in the second dc:identifier field above, the prefix "urn:isbn:" can be added programmatically in MINT or in a simple XSLT to comply with Europeana's example from the ESE version 3.4 specification.

There are no fields in the National Service 4 output format that correspond to europeana:object (full product content will not be linked openly), nor to europeana:isShownAt (since any retail link would privilege a particular retailer, and there is yet no multiple resolution linking service connected to this data service).

In maintaining this neutrality, National Service 4 shows most of the same business model characteristics as National Service 2; there are no usage terms apparent from the API help site, but normally *API users are charged a fee for access to the full (updated) service*. This case again highlights the "service" nature of commercial data providers, as well as the fact that commercial providers are working on state-of-the-art applications of Web data techniques.

Permission to use this data must be negotiated with the database provider and data owner (this is a technical platform only, as discussed in section 4.3.5 and similar in some respects to case study 4.5.8) before any could be provided to Europeana.

One major publisher of general, "trade" books, whose data is included in National Service 4, did respond to the Linked Heritage contributor proposal; this was a significant agreement in princple. Of course, this would take the form of a one-off data upload, rather than a negotiated API access. Unfortunately contact with this publisher was only established at the very end of work on D4.3 so specific details of a prototype contribution could not be discussed.

4.5.10 Large international book data service providers

This commercial service uses several detailed and extensive agreements for both data contributors and data customers. This is obviously due to the need to protect revenue streams which depend on payments to access and reuse the data, as well as to include data for marketing purposes; this income ensures profitability and maintenance of service levels. Both users of the data feeds, and also publishers who wish to list books in any more than the most basic way, pay for these aspects of the service. There is a free element of the service – for basic listings – but enriched data is not free, but is offered for a one-off setup fee plus a smaller annual maintenance fee. This is in contrast to data services run by trade associations, where listing a title (contributing data) is normally free of charge, while enhancements to the data represent value added for the benefit of data users.

This **International Service** charges data *recipients* depending explicitly on the type of data user, the exposure and extent of data, and on a per-title basis for *listing* publishers' books. Library data *users* pay a site licence fee depending on the numbers of sites and users, and other organisational factors.

It is worth noting that annual charges per organisation subscribing to these data services normally range between the thousands and tens of thousands; the only reasons for paying a premium above these (in institutional terms) reasonably affordable rates would be *"for bespoke data files or where the exposure and use of our data puts existing or future revenue at risk"*; this is quoted from a private email correspondence with the International Service, as is the following:

> *"This is a critical consideration for us as all we do… is aggregate data and add value through the curation and enrichment of that data. So exposing our data on a large internet retailer's site or a non-commercial site presents a greater risk of revenue loss (new or existing) than does exposing our data on a niche site.*
>
> *Taken to an extreme, exposing our data to a wider market without commercial gain would be ultimately self-defeating as we'd go out of business and organisations would have to gather data themselves; which isn't that easy…"*

The Service has the following terms for contributors

- Scope:
  - Provision of user's data at the Service's discretion to customers in book trade.
- Duties:
  - Recognition of the Service's rights in the database.

For users of the database, the terms are:

- Scope:
  - Online book retail with the user's own clients.
- Allowed and forbidden uses:
  - Incorporation of the database into the user's own database;
  - Provision of public access to user's database;
  - Internal use for related business purposes;
  - Supply of ISBN, title, price, author, format, and BIC subject codes for third-party search engines, SEO and shopping comparison sites;

- NO redistribution, sale, licensing or third-party provision of public access to the database.

- Duties:
  - Recognition of International Service 1's database and other rights in the database;
  - Display of specified copyright notice along with bibliographic data;
  - Deletion of user's copy of the database on termination of contract.

]

4.6 SUSTAINABILITY

Alongside findings demonstrating the degree technical and commercial viability of contributing data to Europeana *per se*, creating the case studies has made even clearer the investment in time and effort required even to secure a small, one-off contribution of product data. Since contributing data to Europeana is free and relatively simple (or at least well-documented), one can say that if commercial publishers had a strong incentive to do so, they would already have done.

One reason they have not is probably lack of awareness of Europeana and metadata sharing and digital libraries in general, but even once the proposal was discussed with a large variety of publishers, none of them expressed interest in managing the process themselves. Therefore the inherent sustainability of this proposal must be explored, beginning by understanding the difficulty faced by publishers and data services themselves.

The primary reason for this reluctance to engage with Europeana must be seen in the current Creative Commons licensing framework that forces dynamic (and therefore controlled) product data into a static open (uncontrolled) data mould:

- Commercial media companies are generally familiar with the potential “issues” of sharing data openly (as in 3.3.3 and 4.5.10);
    - partly from the now common, analogous experience of dealing with online piracy;
    - and partly from the awareness that successful media product data services often tend in the opposite direction over time, from informal, open collaborations to centrally-maintained, and user fee-supported, commercial businesses (as has been the case with several examples cited in 3.2);
- The most commercial data services cited this explicitly as a reason they could not consider even a minimal data contribution (see 4.4.5 and case study 4.5.10).

Other related reasons are:

- Lack of a clear business justification to devote staff time to ESE data feeds (*e.g.* direct evidence of increased sales potential , or even openly expressed intention of Europeana to *promote* its contributors’ book sales – this is often a stated condition of data reuse, as in case 4.5.6);
- Lack of required in-house messaging standards expertise (even for relatively simple proprietary formats like ESE) and data transformation experience;
    - Either of the commercial contributors themselves, to support data creation specifically for this purpose;
    - Or within Europeana to support use of (and encourage adoption of) existing standards – if Europeana could centrally maintain, over the long term, stable mappings from the major industry data standards to its own formats, this might reduce reluctance;
- The perceived one-off nature of the experiment in the absence of a stated, comprehensive *commercial content development strategy* (either of the Linked Heritage consortium or of Europeana).

In order to produce even very limited, simple test and prototype ingestions, a year of effort from Linked Heritage’s WP4 leader, the WP4 partners and a substantial number of interested co-operators were required. It seems clear that for any long-term partnership, a more solid organisational structure would have to be developed, providing continuity of management, fully justified costs and above all, clear benefits that fit the business plans and organisational objectives of all partners.

In summary, as is clear from the best practice outlined in D4.1 and D4.2, dynamic commercial product data cannot be contributed in one-off project form; it necessarily requires a on-going service basis underpinned by sustainable infrastructure, management and contracts.

4.6.1 Continuity of Management – which partners?

The precise form of general public-private partnership management in terms of organisation, finance, legal and contractual specification has been discussed extensively (*e.g.* in European Commission, 2003) and the rationale and formal arrangements for cultural heritage digitisation partnerships (see 4.4.1) are also well known. As has been noted in section 4.4.1, conditions in the commercial product and data supply chain are very different from those in heritage digitisation, and many existing solutions will not work. Most of the PPP examples in the EC's 2003 *Guidelines* come from environmental infrastructure projects. Nonetheless, some basic points can be made, especially in comparison with existing commercial or non-profit data services and existing cultural heritage aggregators.

4.6.2 Management by public body

The simplest form of partnership would be to provide all aspects of the aggregator service within the public sector; only the maintenance of raw data feeds and management of relationships with the aggregator would be handled by the commercial partner (minimising their costs). The managing entity could be:

- Europeana Foundation itself;
- MICHAEL Culture Association (or similar consortium of heritage organisations);
- Separate, EC-funded service.

In any case, a project-based consortium is not permanent enough to ensure continuity of management in this sense; a company or foundation with long-term political, organisational and financial commitment is required to provide a trustworthy contracting partner and an acceptable level of service.

4.6.3 Management by commercial organisation

The options for direct management within the commercial sector appear far more limited. In principle, commercial content producers (publishers, record labels, film producers, photo libraries) themselves could simply maintain their own data feeds to Europeana without intermediaries. Certainly the practical expertise provided by projects like Linked Heritage should be enough basis for an enterprising company to build their own contribution.

So far, the evidence from the case studies here suggests that no content producers see this type of data exchange as a core activity worth the investment. Those who do count as both content producers and "net exporters" of metadata supporting their products tend to see it as a useful internal asset to be shared only with customers for commercial advantage (*c.f.* case study 4.5.7). Any actors who specialise in data services (*c.f.* case study 4.5.10) strongly control reuse of their services.

Far more promising – for the smaller content producers and markets, at any rate – would be some sort of intermediary with access to infrastructure, dedicated personnel for data enrichment and relationship management, and a greater pool of data sources:

- Commercial content producer or distributor;
- Intermediaries such as:
    - Existing trade organisations;
    - Existing supply chain service providers (often owned by trade organisations);

- Existing cross-sector aggregator(s) (see 6.3).

Even if this could be a feasible option, it must again be stressed that so far, the commercial incentives are apparently neither widely known or strongly convincing or it would be easier to find existing examples in this category.

4.6.4 Data ownership and licensing

Metadata are usually created in the relation between the publisher and the bibliographic agency. The contribution of the two may be different according to the country: publishers may provide an initial record that is checked by the bibliographic agency, possibly corrected and/or enriched and passed back to the publisher; in other occasions the record is directly created by the bibliographic agency, books in hands, with or without an agreement with the publisher; or publishers and bibliographic agencies may be in charge of creating different fields of metadata within the same record. Practices are different country by country, and even in the same country may be different publisher by publisher.

As a result, bibliographic agencies always have *sui generis* rights on the Books-in-print database as a whole and publishers may or may not have rights on subset of the Books-in-print. Then, intermediaries downstream that receive data from the bibliographic agency have no ownership on them, since they simply acquire a license for limited uses that never grant them the capacity to release the data in CC0.

As a conclusion, the contractual constraints indicate that intermediaries such as retailers simply cannot join Europeana under the conditions set in the DEA, while publishers or bibliographic agency may, though this requires an analysis case by case. It is in fact possible that in the agreements between publishers and bibliographic agencies there are exclusivity clauses or other limitations so that neither publishers nor bibliographic agencies can deliver their metadata in CC0 without the agreement of the other.

These considerations are confirmed by the case studies in section 4.5; larger bibliographic agencies, for contractual and legal reasons – and larger publishers also for commercial reasons – are unable to offer even a subset of their data under CC0.

The only exceptions are where the whole data supply chain is collapsed within one organisation, as for smaller publishers (case 4.5.1) and vertically integrated data and content aggregator-retail platforms (cases in 4.5.9), or where a data service provides purely technical services while individual content releasers retain full rights in their data (case 4.5.3).

Unfortunately none of these options appear to scale in terms of data quantity beyond a local level requiring high individual maintenance (as in case 4.5.1), nor to offer the richest data in terms of descriptive text and previews (see studies “National Service 1A and 1B” in case 4.5.9).

5 POTENTIAL INCENTIVES FOR COMMERCIAL DATA PROVIDERS?

In principle, providers could send metadata records describing *items of content* more granular than a discrete product (e.g. a book cover), or for individual *retail offers* (e.g. using a retailer's local ID for a specific product at a specific price, with specified delivery conditions, etc.) – however, following industry best practice we have focussed on identifiers and metadata records *for products.* Any other focus is likely to involve duplicating work, either for Europeana or for the users of their data – whether in the portal or as LOD downloads. Bearing this in mind, we can analyse the likely benefits for the various types of potential partners.

Unless contractually obliged not to, as described in 4.6.4, in principle every operator in the e-book or metadata value chains can join Europeana. An organisation joining Europeana is asked to provide:

1. One metadata record[112] for every e-book;
2. One "preview" associated with every record (e.g. the thumbnail of the cover image);
3. One (and not more than one) URI that resolves in a persistent[113] way to the content.

Once again, we will take the e-book supply chain as illustrative of the general problems, since it is well developed, and breaks down into a full range of possible roles. Of course, it would be possible to add further players that contribute *content* to this supply chain (*e.g.* authors and editors themselves, as well as picture and textual archives) but it would be unusual that these would also contribute new *metadata* to the respective value chain.

The following table resume pros and cons for every actor in the chain that might provide data to Europeana.

	Pros	Cons
Publishers	• Control the sales rights in the e-book and therefore the largest interest in increasing visibility; • May identify the best URI (or URIs) to address users	• Huge numbers of individual publishers, which increases the management cost for Europeana unless there is an aggregator of their metadata; • May not have the best metadata available along the chain • May not be able to purchase the product from the publisher's URI, as many publishers do not have the ability to fulfill products direct to consumers
Distributors	• Sometimes have much better metadata than publishers; • Already act for a number of publishers (ie. the data is aggregated); • May act on behalf of publishers to provide the most qualified URI(s)	• Often do not have the right to re-use the metadata for the purpose; • Metadata is not in their core business so may not be able to provide standard output formats

[112] To an extremely minimal specification; see 4.2.1.
[113] Persistence is a key term and must be defined carefully in a commercial context; a concise, thorough discussion can be found at http://www.doi.org/doi_handbook/6_Policies.html#6.5

	Pros	Cons
Retailers	• Strong interest in increasing visibility for their offering; • Generally have very good metadata (though rarely have the rights to allow its reuse or redistribution)	• Interest is limited to attracting Europeana users to their own retail website. Any preference or support for a single retail outlet would be seen as unfair; • Alternatively, if Europeana accepts all retailers that wish to join, users will receive multiple records for the same book; • Otherwise, Europeana would need to develop an identifier-based methodology to de-duplicate the records and present multiple URLs leading to retail websites
Data Aggregators	• Have very good metadata; • Often also hold the necessary rights to reuse it	• Do not always hold URLs pointing to retail websites; • The "multiple records for a single product" issue that applies to Retailers also applies here
Bibliographic agencies	• Have by definition the best metadata in the supply chain	• No direct interest in providing the metadata for free to Europeana; • An analysis of their business model is required to understand possible value for them to join; • Usually do not have a single URI resolving to the content or product

Since bibliographic agencies specialise in metadata and identifier management, they have the most data records and highest quality metadata, and in principle – where multiple suitable retail URIs *are* available – could avoid the commercial fairness barriers to contributing retail offers, their business model is discussed (in 5.1.1) in detail.

5.1.1 Existing business models of bibliographic agencies

This synthetic analysis seems to indicate that bibliographic agencies (also known as 'books in print] agencies, since they aim to provide an authoritative list of commercially available products) are the most qualified actors to provide high quality metadata to Europeana, if they act *on behalf* of publishers. The objective is to combine the largest interest in the value that Europeana can offer in terms of visibility – for publishers – and the agencies' specialisation in dealing with high quality metadata. This is closest to the model in place in Europeana where metadata aggregators offer a single access point for multiple content providers.

The key issue remains the commercial interest for bibliographic agencies to join Europeana. This depends on their business model. The revenue sources for books-in-print are essentially the following:

a) Publishers pay to be listed in the books in print database. Bibliographic agencies offer value to publishers precisely in the dissemination of book metadata, which increases the market potential for every book;

b) Access to data is sold to professional users, such as booksellers, **libraries** or publishers. This model was originally based on selling printed catalogue of in-print titles, then moved to CD-Rom, and is now based on web platforms that users subscribe to access. **There are no known cases where such subscription is also offered to the end consumer market**;
c) The whole database is licensed to users for integration with their own IT systems so as to provide **services to end users**. Typically, this is the case of Internet **bookshops** that use the database to allow users to search among in print titles and to reach rich information about them. This is also a clear incentive for bibliographic agencies to enrich metadata records with elements that in Europeana's terms may be called "previews", such as the cover image, abstract, table of content, excerpts, *etc.*;
d) On top of the database some companies also created further value added services, like teleordering systems (routing electronic orders from retailer back to publisher or publisher's distributor) or the provision of sales statistics.

Integration of commercial data into Europeana's portal would thus combine aspects of b) and c), as Europeana would become a kind of "library" with links through to a virtual "bookshop" with offers from many retailers. However, the release of open data to end-users without charge is a stark difference from the existing model in b).

The combination between the different sources of revenue varies in different European countries, depending on the history of the individual companies – decisions made decades ago influence the current equilibrium – and on the size and maturity of the book market that every bibliographic agency serves. For example:

- In Electre (the French BIP) or in Alice (the Italian one), publishers' titles are listed for free and all the revenues for the Cercle de la Librerie and Informazioni Editoriali respectively come from other sources (ie. from resale of the data);
- In the UK, Nielsen lists all titles for free, but charges publishers for more comprehensive or enriched listings, and derives further revenues from bulk data sales and value-added teleordering services;
- In relation to the added-value services, in France teleordering services are provided by a different company (Dilicom), whilst in the UK or in Italy this represents a key element of the value proposition of the bibliographic agency;
- Sales statistics, on the other side, are very important for Nielsen in the UK – which provides similar services in other countries – while in Italy, Informazioni Editoriali produces the raw, row-level sales data from the teleordering system and sells the raw data to other companies that compile statistical analyses for publishers.

As a common trend, the second revenue source (selling access to data) is declining, since professionals can reach very similar information from other sources, while the third, **licensing the database as a whole, is growing with the explosion of e-commerce and other web based services.**

Finally, it is important to note that in several European countries, books-in-print databases simply do not exist, though initiatives are in progress to create them[114]. Because of the commercial and cultural importance of books in print systems, and to cope with the difficulties of establishing such

[114] The most relevant is within the ARROW Plus project (see http://www.arrow-net.eu). The creation of books in print databases is planned in a number of countries (namely Poland, Portugal, Bulgaria, Hungary, Latvia, and Lithuania). Currently, the focus is on establishing a long term business plan to ensure sustainability of the initiatives. In this, difficulties remain in finding the right combination between the revenue sources, and a sustainable cost structure. Other initiatives have been launched thanks to the support of the CCEBP (Funds for Central and East European Book Projects – see http://www.ceebp.org).

resource for the book trade in countries where market forces alone do not allow their sustainability, many recommend that such initiatives should receive public support. European support would facilitate interoperability between books in print, facilitate cross-border discoverability, and create incentives to share costs among multiple BiP services, which may be a crucial factor for reaching the break-even point.

It is crucial to note at this point that any operation actually delivering commercial product data to Europeana, even if that data is minimal, or if the operation is based within a larger organisation, will inevitably replicate some of the work of a books in print database, since it will necessarily include elements of relationship management, data quality and currency maintenance and standards compliance. We stress this point here to clarify that much of this work is already being done, sometimes, as in the case of ARROW PLUS[115], funded directly by the EC, and at a minimum, any new services or data provided by Europeana should not negatively impact existing books in print services.

5.1.2 Incentives for contributing to Europeana for bibliographic agencies?

Starting from the realities of the books-in-print market, in theory the incentive for a bibliographic agency to join Europeana would be maximum when all revenues come from the first source (publishers pay to be listed) and would be zero if no revenues come from this source. The reason is evident: the benefit in joining Europeana is entirely for the content providers, *i.e.* the publishers, who increase visibility for their titles, and thus can expect to increase sales.

If publishers are the only clients of the bibliographic agency, contributing to Europeana will increase the value offered to clients. On a theoretical level, in this case the decision making process is very simple: bibliographic agencies compare the cost of contributing with the value perceived by publishers. If the benefits are higher than the costs they join. Under this hypothesis, promoting Europeana in the e-book environment should be simple:

- *reducing costs* on the one side, primarily through projects like Linked Heritage;
- and *increasing awareness* of the value of Europeana on the other side.

In the opposite circumstance, when titles are listed in the database for free, there are no incentives for the bibliographic agencies to contribute; not just because there is no the possibility to pass on value to clients. The more serious obstacle is that providing data to Europeana without charging for it may jeopardise the business:

1. How can the bibliographic agency justify provision of data for free in one case, while other customers (very often including public sector organisations) pay for the same data?
2. How can the significant costs of collating, enriching and updating the database, and maintaining the feed to Europeana, be met without passing them on directly to Europeana?

In reality, bibliographic agencies have a mixed business model that does not coincide with the extreme hypotheses here described. This tends to weigh against provision of useful data at no charge; the reality is that rich data is already highly in demand; Europeana's plan to redistribute it as linked open data does not arrive on an empty playing field, but rather a highly complex, crowded and competitive one.

On the other hand, even when titles are listed for free in the database, publishers may still be seen as "clients", since the quality and the timing of the data received by publishers are crucial from the cost side and in return for this consideration, again, any books-in-print operator has an interest in providing at least some value to publishers. In these cases, when the service is likely

[115] See http://www.arrow-net.eu/what-arrow-plus

subsidised by other, profit-making activities, public funds or other cultural foundations, there may be justification for contribution of a minimal subset to Europeana.

5.1.3 The use of ISBN-A to provide multiple links via a single URI?

In the context of large-scale Internet book search, some publishers have shown concern over controlling links associated with their own books. One solution has been proposed combining existing standard identification and resolution systems: the ISBN as unique identifier for books, already used to create links to different retailers in automated ways; and DOI supporting multiple resolution, *i.e.* the capacity to "return as output of several pieces of current information related to a DOI-identified entity – specifically at least one URL plus other defined data structures providing additional information"[116].

The ISBN-A is "an 'actionable ISBN'; an existing ISBN is enfolded in a DOI"[117], providing a service either to a person pointing a browser at the DOI resolver, or an automated agent requesting metadata for reuse. Individual ISBN-As need to be registered separately from the ISBNs. ISBN-As are not created automatically for all ISBNs, nor are they necessarily free of charge[118].

An integration of the ISBN syntax into the DOI syntax means that for every ISBN it is possible to create a DOI that incorporates it. At that point, all the features of the DOI can be exploited, and in particular, the resolution service. Like any DOI, an ISBN-A can be provided in form of a URI[119] or live Web link.

At present this tool is used to separate the management of the links that appear in a webpage and the management of that webpage. A "landing page" for each product, maintained by the ISBN-A service, is displayed when the DOI link is clicked; here, in a standard format based on ONIX for Books, key product information is offered to the user, including any number of further links, with some information describing those links' targets. This usually includes a number of different retailers' offers.

In Linked Heritage's context, if a metadata provider registers in Europeana a URI based on ISBN-A, this will allow the publisher, who is authorised to manage the URLs associated with each ISBN-A:

1. to direct users to multiple resources of their choice. This may include, for example:
    a. multiple services to buy the e-book or print book;
    b. different information sources such the author profile, the author's blog (if any), reviews that appeared on prestigious websites, richer metadata, including abstract, table of contents,*etc.;*
2. potentially, to provide "behind-the-scenes" rights information for downstream users of information republished by Europeana;
3. to change or upgrade the landing pages, and potentially extra metadata services, over time without any change in the Europeana system.

[116] See: http://www.doi.org/doi_handbook/3_Resolution.html
[117] See: http://rdtf.jiscinvolve.org/wp/files/2009/09/rdtf-hazel-woodward-cranfield.doc
[118] But DOI is intended to enable cost-recovery prices, and such would be the aim of a heritage sector ageny; *c.f.* http://ispiders.blogspot.co.uk/2007/10/buying-selling-doisand-same-for.html

[119] Cf. *ISBN and ISBN-A*, in DOI Fact sheets: www.doi.org/factsheets/ISBN-A.html. For example, starting from the ISBN: 978-88-89637-41-8, it is possible to create the ISBN-A (a DOI): 10.978.8889637/418, and then the URI: http://dx.doi.org/10.978.8889637/418.

The ISBN-A could potentially become a community-led tool, similar to others characterising Web 2.0, and the recent linked data enhancements provided by DOI implementations DataCite and CrossRef[120].

[120] See, *e.g.* http://bitwacker.com/2010/01/19/the-doi-datacite-and-linked-data-made-for-each-other/ and http://www.crossref.org/CrossTech/2011/04/content_negotiation_for_crossr.html (as well as http://www.crossref.org/CrossTech/2008/05/metadata_reuse_policies.html)

6 POTENTIAL REVENUE STREAMS

Given that a desirable outcome of this groundwork is large-scale aggregation of commercial data to Europeana, and that this will almost certainly reach beyond what Europeana Foundation alone can support, it is important to consider possible external funding streams.

6.1 PUBLIC FUNDING

It is assumed that a service-scale commercial data aggregator would run on an EU-wide or at least multi-Member State basis. Therefore some aspects of the rationale behind other EU-wide data service projects might apply (see ARROW Consortium, 2011):

- Public project funding with continued support:
    - Supporting EU cultural policy (in this case, through enrichment of Europeana and offering of new services to the European public);
    - Supporting publically-funded institutions;
    - Political declarations and action plans (in this case, the *Comité des Sages*' report, mandating "digitisation and online accessibility of in-copyright material").

In analogy with the suggested budget item "diligent search" in the ARROW Business Model, one might suggest a "promotion of cultural commerce" item for national cultural enterprises and projects; of course, in contrast to the legal requirement for diligent search, this would need to be a positive item attracting grant funding upon successfully demonstrating capacity to add value through use of the commercial data aggregator.

6.2 PRIVATE FUNDING

Again, based on the funding possibilities for books in print services, another option would be:

- Direct or private funding:
    - Payment for use (in this case, payment for adding a data feed to the service);
    - Payment in kind (in this case, users of the service are unlikely to also be data providers, but a data service provider might consider hosting the service and thus receive privileged terms to add its own clients' data feeds).

Direct funding could possible include subsumption of "the service" as one (none-core) activity among others for a books-in-print (or equivalent) data service, which would already aggregate many of the required elements (especially commercial availability). Some suggested funding sources for books-in-print (ARROW Consortium, 2012) might also apply to an aggregator:

- Publishers pay for enhanced title entries *e.g.* long descriptions, reviews, author interviews *etc.*;
    - This in fact mirrors the business model of some commercial data aggregation, enhancement and redistribution services such as that provided to the book trade by Nielsen. However, the linkage of the service to demonstrable uplifts in sales, or other benefits must be clear;
- Publishers and other trade organisations buy advertising space on the aggregator site (though unless this site attracts either a large volume of traffic from consumer web users, or a reasonable volume of highly-targeted traffic from business web users, the value of that advertising would likely not be sufficient to support the provision of the service);
- Publishers pay for the facility of sending data to a single hub for conversion to the data formats of the major booksellers, wholesalers etc. who require their own data formats;

- This mirrors the business model of some commercial book data service companies (*e.g.* Firebrand or Booksonix, whos service includes both elements of 'software as a service' and data conversion and redistribution).

Also along the lines above might be collection of affiliate fees from Internet retailers to have their page for the products added to the metadata.

The primary source of funding recommended here for new books-in-print services, in countries which do not already have them, is direct European funding for the acknowledged societal good they provide, perhaps with an explicit condition of providing a minimal subset of data to Europeana under favourable terms, while retaining the right to provide richer data to paying commercial customers at a higher rate, in order to remain cost-effective and operational in the long term.

6.3 ALTERNATIVE MODELS REDUCING COST / REVENUE STREAMS

Various motivations for organisations to contribute their data (or facilitate contributions from others) can be identified which do not depend on either direct commercial incentive or direct EU heritage policy:

Motivation	Funding structure	Examples
Providing legal alternatives to pirated content	Not-for-profit, grant funded, advertising and referral fees	Findanyfilm.com Moviepilot.de
Promoting national cultural heritage and cultural industry	Not-for-profit, direct revenues from content sales (also advertising in the case of Ina.fr)	Ximon.nl Ina.fr

The central question that remains in each of these cases is: could the revenue from associated activities fully support provision of an EU-wide data service, or would it require external subvention[121]?

Existing platforms constitute in some sense an alternative to Europeana on a national level. Important examples at this level are Enclave, and also France's Gallica 2 (a Europeana contributor). All of these services offer portal search (in the case of Enclave, potentially also full-text search via DILVE, and Gallica 2 offers limited full-text preview) but restrict reuse of their metadata. In this sense they offer access to richer data than Europeana.

However, neither the national public-private portals, Gallica and Enclave, nor Europeana itself, rely on advertising, referral fees or direct sales for income; none of these revenue models fit with Europeana's current service structure. To adopt any of these, Europeana (or its contributing aggregator) would become a competitor with existing, more commercial services. A notable exception, is Ina.fr which is both a Europeana contributor and a semi-commercial service, selling commercial licences for its content and receiving revenue from advertisers. This model could certainly be adopted by a new, third-party portal contributing a minimal subset to Europeana.

The models in the table above follow the example of existing commercial collaborations established by special, usually national, interest groups, and it is possible to imagine future

[121] Analogous attempts to create central aggregators of cultural heritage *content*, such as AMICO, without central subvention, have raised similar questions; see, for example: http://www.archimuse.com/publishing/ichim03/012C.pdf

services working in parallel in other countries and industries. Whether such models can scale to a Europe-wide level is a far wider question.

Industry sector	Possible supporters (illustrative only)	Relevant programmes
Books	FEP/FEE (Federation of European Publishers)	European writing awards
Film & TV	FIAD (Federation of European Film Distributors)	Film awards and promotion
Recorded music	IFPI (International Federation of the Phonographic Industry)	Anti-piracy
Photography	CEPIC	Promotion of metadata and licensing information best practice

Of course, a mixed consortium of these organisations, for whom a central data adaptor and Europeana aggregator would be of value, but whose reasons for doing so might vary, could jointly support an EU-wide service at a shared (and thus probably lower) cost.

7 POTENTIAL COSTS

Because aggregating commercial data is so similar to the work of books-in-print services, and because if it is not done independently, many of the costs will be met by books-in-print by default, these costs have been adapted from those identified for books-in-print by the ARROW Consortium (2012):

- Hosting costs;
- Hardware and software maintenance:
    - Technical support and upgrades for hardware and software;
    - Maintenance of mappings between industry standards;
    - Maintenance of Europeana upload format mapping;
- Staff:
    - Management;
    - Editorial:
        - Quality checking;
        - Mapping adjustment;
        - Assurance of currency and deduplication;
    - Relationship management;
- Marketing and project management of new feeds or services.

The **absolute requirement of ongoing maintenance by staff with relevant expertise** has been confirmed by the experience of managing test and prototype data within the lifetime of Linked Heritage.

8 CONCLUSIONS AND RECOMMENDATIONS

Based on the theoretical considerations, business model scenario mapping and real cases studied, Linked Heritage is in a position to make recommendations for realising public-private partnerships for contributing commercial data to Europeana.

8.1 FOR EUROPEANA

In order to attract substantial commercial data contributions, it is recommended that Europeana should consider the following three enhancements, to level the balance between the value delivered to Europeana (data richness) and to the data contributor (ease of discovering and purchasing relevant products).

Europeana's possible actions are listed here *in order of increasing investment required*, with corresponding returns in terms of the amounts and richness of commercial data available, and the type of services that could be developed.

Licensing services for Digital Objects, as already noted in 3.5, are in reality a more urgent requirement; these are not discussed here in detail, but the new directions taken in the Europeana Creative[122] and Europeana Cloud[123] projects are welcomed. One aspect of licensing systems that does not appear to have been accounted for in either project is the creation, maintenance and resolution of identifiers for content, parties, rights and transactions. This problem is touched on here in recommendation 8.1.3, and also in the *D4.1 Addendum*.

8.1.1 Dedicate new resources exclusively to acquiring and maintaining commercial data

The contribution of even small amounts of commercial data will require, at minimum, dedicated staff and resources budgets comparable to those used to achieve prototype data contributions in Linked Heritage. Due to the specialised nature of commercial products, data formats and services, Europeana will need to consider recruiting staff with this expertise, and not just relying on PPP advisors from within the heritage sector itself.

Even in cases such as small content companies (case 4.5.1), or data services able to liaise with their clients (case 4.5.8) who could then directly grant usage rights over data, where a direct feed or one-off contribution of a whole title list may be commercially, legally and technically feasible, **significant amounts of work are required to achieve delivery of the data, and this will not be funded by small contributors themselves**.

It must be stressed that although the documentation produced by Linked Heritage WP4 is a solid basis for this work, each case will present its own challenges and will require additional research, development and management of data and relationships with suppliers, and so cannot simply be added to the duties of existing personnel at Europeana Foundation's offices.

Commercial data specialist(s) could be recruited on a project, or ideally, permanent basis and would be responsible for:

- Identification of potential commercial content to be delivered to Europeana;
- Liaison with potential content providers at all stages of the supply chain, and networking to build awareness of Europeana within commercial sector;

[122] See http://pro.europeana.eu/web/europeana-creative
[123] See http://pro.europeana.eu/web/europeana-ecloud

- Negotiation of agreements with commercial content providers;
- Providing feedback to Europeana Foundation on technical, legal and commercial requirements of commercial content providers, and contributing to project proposals to enable delivery of data through new or existing services;
- Technical support to commercial contributors, including:
    - Assistance with producing and maintaining standard data outputs if not already supported;
    - Filtering of existing data for relevance and commercial / legal acceptability;
    - Transformation of commercial data into heritage formats, including Europeana's import schemas;
    - Upload and update of files;
- Research and dissemination of commercial sector data formats expertise within Europeana network and beyond, to support future contributions and PPP projects.

From the wide scope of this brief, and the potentially vast amounts of data to be exploited[124] it should be clear that this is full time work, perhaps with a division of responsibilities (*e.g.* roughly divided among management, research and communications, technical implementation, and commercial relationship management) among a small team within Europeana Foundation.

8.1.2 Use separate licensing frameworks for basic and rich metadata

A first step towards data enrichment would be to allow a more nuanced licensing framework for data points beyond "factual" data such as those enumerated in the "proposal accepted" case study in 4.5.9; which was the most "commercial" service that agreed in practice to send book data to Europeana.

It is worth noting the similarities and differences between the data supplied as "public domain" in this case and those used in an ISBN registration, a minimal subset of bibliographic data used primarily to distinguish between otherwise similar products, and CIP (Cataloguing-in-publication) date, used to offer a minimum starting point for libraries' catalogue records (which are usually made available somehow to the public):

Data field	ISBN registration message[125]	CIP[126]	Case study 4.5.9
ISBN	Y	Y	Y
Product form	Y	(included with ISBN)	Y
Title	Y	Y	Y
Series	Y	Y	N
Contributor	Y	Y	(author only)
Edition	Y	Y	N

[124] See also Linked Heritage D4.2 (http://www.linkedheritage.org/getFile.php?id=394) for estimates of the amounts of commercial data in each sector.
[125] See ISBN user manual: www.isbn-international.org/pages/media/Usermanuals/ISBN Manual 2012 -corr.pdf
[126] Combined hypothetical CIP format taken from British Library (http://www.bl.uk/bibliographic/pdfs/cip.pdf) and Library of Congress (http://cip.loc.gov/cipman/) examples; specific fields may not appear in both UK and USA CIP programme outputs

Data field	ISBN registration message[125]	CIP[126]	Case study 4.5.9
Language(s) of text	Y	Y	Y
Imprint	Y	Y	N (but may be the same name as "Publisher")
Publisher	Y	Y	Y
Country of publication	Y	N	N
Publication date	Y	Y	Y
Retail site URL	N	N	Y
Subject classification	N	Y	Y
Table of contents	N	N	N
Extents / measurements	N	Y	Y
Notes	N	Y	N
Control numbers, national bibliography numbers	(ONIX registration message requires a local "RecordReference" ID)	Y	N
Record source	Y	Y	(included in all ESE data)
Price	N	Y	N

No subject classifications would appear in product registration messages, such as those for ISBN, as they are not always strictly factual and depend a certain amount on the discretion of the classifier and on local practice (see comments on this in section 4.4.12 and case study 4.5.9).

On the other hand, many other "factual" data about the product could be added which would serve to add some limited marketing value; edition number or type; related works and products, e-book usage constraints[127]. Analogous "facts" about other media could be inferred semi-automatically or found directly in the relevant identifier registration standards. Prices are not useful to the general public unless kept up to date; but see below for the relevance of books in print services.

CIP programmes have a strong relationship to books in print (BIP) services in many countries, where BIP is not profitable as a commercial operation, and the general database of commercially available titles (BIP) can be produced only by basing it on a centrally funded bibliography, such

[127] See the ONIX for ISBN Registration message outline for full details: http://www.editeur.org/118/ONIX-ISBN-Registration-format/. For an overview of other identifier registration formats, see Linked Heritage D4.1: http://www.linkedheritage.org/getFile.php?id=283

as the national bibliography maintained by many national libraries. The final link in this chain is the fact that often a national library will be the ISBN registration agency for such countries.

Even given that one data service provider was willing to contribute the ten data fields from case 4.5.9 on a CC0 basis, and no other large datasets could be found to contribute despite exhaustive searching, it is clear that **more flexible data licensing is the only way to move beyond this impasse**. Otherwise, richer data will remain "locked" in mid-level aggregators (like those managed by most Europeana Network projects) or paid-for services like CIP and books in print, who require, at minimum, some explicit agreement on usage terms and conditions in order to protect their business model.

8.1.3 Support multiple resolution of unique product identifiers

In parallel to enabling separate licensing and reuse terms for commercial data, a reasonably cost-effective improvement would be to implement, or at least enable interaction with managed, resolvable unique identifier systems like ISBN-A[128] and EIDR[129] (see also 4.4.14 – these services curate data that is relatively "open").

In this way, using multiple resolution, the problem of multiple retail offers for each unique product (see sections 4.5.9 and 5.1.3) could – at least potentially – be overcome in a sustainable way.

As with the use of separate licensing schemes for product data, this option in principle should require no extra commitment of resources, since it would be based on existing industry standards, best practices and technology; in reality it would probably require substantial development and perhaps a future project to put in place (Europeana might also consider supporting multiple resolution of identifiers for cultural heritage objects and heritage institutions' digital objects, which would be a very ambitious but potentially useful project, especially for picture licensing); hence these options appear further up the scale of expense than simply employing a team of specialists to manage publisher relationships using the existing tools.

8.1.4 Enable separate search of textual "preview" material

Building on the licensing separation described in 8.1.2, richer material could then be included to support more effective search. If this rich data were hosted centrally at Europeana's server[130] it would have to be under an explicit guarantee that it would not be included in APIs or Linked Data sets that are released under CC0.

One alternative possibility might be to formally define this type of textual data as part of the "preview" mentioned, but not defined in detail, by the existing Data Exchange Agreement[131]. In the DEA, the term "Preview" in fact includes "text files", and the URLs to these "Previews" can be excluded from data exports; but it is not explained how such textual previews might be displayed in the portal nor whether they can be included in the search function. It must be assumed, in the absence of any other mentions that previews other than still images simply cannot yet be contributed to Europeana.

[128] See http://www.isbn.it/ISBNA.aspx
[129] See http://www.eidr.org/
[130] For example, using the europeana:unstored data element that seems to have been originally intended for data used only for search, but not display; see http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57 - but note that use of europeana:unstored is NOT recommended for commercial contributors due to uncertainty as to its technical use and contractual status
[131] See the DEA outline at: http://pro.europeana.eu/data-exchange-agreement and the full legal agreement at: http://pro.europeana.eu/documents/900548/8a403108-7050-407e-bd00-141c20082afd

A revised (but not substantially different, *i.e.* "backwards compatible") DEA that clarified the rights status and agreed uses of specific pieces of marked-up preview text submitted within metadata batches to Europeana could offer a simpler stating point.

One technical enhancement might be to enable submission of search queries from Europeana to a separate "preview hosting service", that would deal with textual preview material such as long product descriptions, contributor descriptions, reviews, criticism and extracts from products. The hosting service would then return actual "results" (identifiable products), perhaps together with short extracts from the full content (which would remain inaccessible to both Europeana and the search user). A similar service is possible already within platforms such as DILVE[132].

Whether this hosting service might exist as a pure search repository or a mid-level aggregator with an open Web portal for access to but not redistribution of its content, is another possibility to consider when exploring funding options, as such rich content, protected by adequate licensing, might attract cultural subventions, advertising or user subscriptions.

8.1.5 Support EC funding for set-up of aggregators or data adaptors for in-copyright materials

Provision of a commercial product data aggregator for Europeana could be a common goal with other initiatives of the EC, chiefly ARROW Plus. The EC's commitment to enriching Europeana's content could be a justification for supporting one or more services that aggregate book and other product data in European countries where creating sustainable services has thus far proved difficult, while feeding data to Europeana could be made condition of funding in their contractual arrangements, at least for an initial period.

As a partner on such service set-up projects Europeana could provide several types of support in in return for benefiting from direct influence over enrichment of its data; for instance:

- Quality control of data and specification of which fields will be most useful to Europeana's portal and linked data services;
- Marketing and promotion of start-up services;
- Assistance with networking and cooperation across commercial and heritage sectors (*e.g.* libraries and book publishers; film and TV archives with broadcasters and studios);
- Finding opportunities for future project funding to expand services and make them sustainable.

8.2 FOR THE LINKED HERITAGE CONSORTIUM

The Linked Heritage consortium has supported this work on PPP, and already possesses much of the expertise required to support aggregation of commercial product data for contribution to Europeana.

However, WP4 has shown that under the current conditions, such aggregation is not scalable beyond the "prototype" level, and even if more robust licensing terms are introduced in future, Linked Heritage cannot support commercial data services without introducing sustainable management structures and contractual agreements that guarantee security of the data.

On this basis Linked Heritage partners could consider providing aggregation services in partnership with Europeana using MINT, the LIDO format and possibly some form of identifier management system developed for this purpose or adapted from ISBN-A.

[132] See http://www.editeur.org/files/Events pdfs/Supply chain 2010/ Peraita_iscs_frankfurt2010_dilve_v02_nt.pdf

9 GLOSSARY OF TERMS

Most of the terms below are defined as in the Commercial Data Contributor Agreement, for ease of reference throughout this report; these appear below in *italics* and have definitions in "quotes" with no external attribution. For other technical terms related to the design and use of metadata, readers may refer to Linked Heritage deliverables D4.1 and D4.2.

Term	Synonyms	Definition
Record	Data item, data record, product record	"An XML document containing metadata, part of the Data Contribution, describing one, and only one, Commercial Product"
Data element	Field	The constituent parts of a Record (see above) – specific XML elements or attributes[133] - expressing one pre-defined category of information (see D4.2 for discussion of information categories[134])
Data set		A particular, identifiable collection of Records; in this report, it mostly refers to Contributions and Publication Subsets
Service	Data service, information service, ["database" in some contexts of ongoing data provision]	In distinction to a one-off Contribution; an ongoing provision of data in response to ad-hoc requests, or as provided for in a prior agreement (*e.g.* monthly updates including any Records newly created in the preceding month)
DoW	Description of Work	The contractual basis of an EC project such as Linked Heritage
Test Data		"A Data Contribution for which the Contributor has specified that no Publication Subset shall be passed to Europeana. Test Data will only be used internally to Linked Heritage."

[133] For basic details of XML elements and attributes, see http://www.w3.org/standards/xml/core
[134] See also http://www.jiscdigitalmedia.ac.uk/guide/an-introduction-to-metadata for an example of this term in use

Term	Synonyms	Definition
Prototype Data		"A Data Contribution for which the Contributor has specified a Publication Subset to be passed to Europeana by EDItEUR."
Offering	Offer, retail offer	A commercial product in a retail context; this is the "offer" to buy the product, so normally comprises both information *about the product* and information about *how to obtain it*, price, supplier, territorial availability *etc.*
Content	Intellectual content, creative content, copyright content (also database right content, although rarely stated thus)	The informational or symbolic-conceptual aspect of media products; that which is protected by copyright or related intellectual property rights (especially database right in this report), broadly similar across the EU with some national nuances
Data Contribution		"A set of metadata records in XML format made available to EDItEUR under the terms set out in the Commercial Data Contribution Authorisation Agreement and Commercial Data Contributor agreement appended thereto."
Contributor		"Commercial organisation who passes Data Contributions to EDItEUR for use within Linked Heritage under the terms specified in the Commercial Data Contributor Agreement."
Publication Subset		"The subset of data elements of each Record within the Data Contribution that, by under a Commercial Data Contributor Agreement, EDItEUR passes to Europeana."
Digital Image File		"The highest resolution digital image file representing a Commercial Product; normally (but not always) found at the Retail Web Page for that Commercial Product"
Thumbnail Previews		"Low resolution previews generated by Europeana from any available Digital Image File for use in its www.europeana.eu portal, and hosted at Europeana's own servers."

Term	Synonyms	Definition
Commercial Product		"A commercially available, copyrighted creation available for retail purchase in a specific form from the Contributor and described by a Record within the Data Contribution."
Retail Web Page		"A Web page, whose URL is provided in the Record for a Commercial Product, where that Commercial Product can be bought via ecommerce by the general public."
DEA	Data Exchange Agreement	"The terms and conditions under which EDItEUR will pass the Publication Subset of Data Contributions to Europeana. This gives Europeana the right to re-publish the Publication Subset under the Creative Commons CC0 framework."
Cooperation agreement	Linked Heritage agreement	The terms and conditions used to enable organisations outside the Linked Heritage partners' consortium to contribute, whether through data uploads or by sharing expertise and dissemination effort
Intrinsic metadata		Data about a product that is *typically* either a) copied from part of the product's content (and usually added at the source of the product; *e.g.* title, creator, date of publication) or b) a basic, practically indisputable categorisation of the product, *e.g.* product forms like "print book" or "e-book", a published, centrally-managed product identifier, or a special feature or measurement like page numbers, running time, or DVD extras[135]
Extrinsic metadata		Data about a product that is *typically* a) added after publication / release of the product and b) in some way subjective, open to alternative statements, or at least normally added by some other agency than the product's producer (*e.g.* library subject classifications added after publication of a book; a quote from a movie review added after release)

[135] Although not specifically addressed in this report, "intrinsic" metadata could also include detailed, automatically-generated file information, especially for digital photos. See *e.g.* http://my.safaribooksonline.com/book/operating-systems-and-server-administration/microsoft-sharepoint/9781118283646/chapter-25-content-types-and-metadata/a2_33_9781118273814_ch25_html or http://www.w3.org/2007/08/video/positions/Schepers.html

Term	Synonyms	Definition
Partnership	Agreement	Formal, bi-lateral exchange of explicit commitments
Commercial [sector]	Private sector (in contrast to public-funded)	Business sectors at least partly based on mutual exchange of goods and services, normally, but not only, for financial return and some profit
Exposure		The likely audience / readership of the data to be exchanged; in this context, mostly equivalent to “users of Europeana”
Coverage		The specific group of Commercial Products described by a Data Contribution; this may not be the whole possible list of Products available from a commercial media company, for example, because some are deemed less relevant to a particular group of users
Currency		How up-to-date the information in a given Data Contribution is; given that it is accurate, how far it represents the most recently available Products
Template		The selection of data elements that make a Record within a Data Contribution
Product form		A basic product classification with varying degrees of emphasis on medium, format, encoding and usage constraints, *e.g.* book, e-book, paperback, DVD, streaming video, picture reuse licence
Business model		Organisational and financial arrangements for providing a sustainable service
Business case		Analysis of costs and benefits providing a rationale for an organisation to provide a service

10 REFERENCES CITED IN THIS REPORT

Adams, A. and Davenport, J. (2004). *Copyright Issues for MKM.* In: Mathematical Knowledge Management. Springer , Berlin / Heidelberg. pp.1-16. http://dx.doi.org/10.1007/978-3-540-27818-4_1

Julie Allinson, J. 2011. *OpenART and Open Licensing.* http://yorkdl.wordpress.com/2011/07/15/openart-and-open-licensing/

ARROW Consortium. (2011). *ARROW Business Model.* Online: http://www.arrow-net.eu/sites/default/files/ARROW_Business_Model.pdf

ARROW Consortium. (2012). *Business Models for Sustainability.* Unpublished discussion paper.

Attanasio P, 2008. "Tecnologie standard per la 'biodiversità' editoriale", presentation at the conference "EdiTech 2008 - Editoria e innovazione tecnologica", Milano, 27 giugno 2008, slides available at http://www.editech.info/leftmenu2008/Download.aspx.

Alex Ball, A. 2011. *How to License Research Data*. http://www.dcc.ac.uk/resources/how-guides/license-research-data Digital Curation Centre

Bekiari, C., Doerr, M. and Le Bœuf, P. (2010). FRBR object-oriented definition and mapping to FRBRer (Version 1.0.2) Available at http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.2.pdf

Bérard, R. (2011). Free library data? [online]. http://liber.library.uu.nl/index.php/lq/article/view/7997 http://liber.library.uu.nl/publish/issues/2010-3_4/index.html?000512 [accessed July 2011]. *Liber Quarterly.* 20(3/4).

Berne Convention for the Protection of Literary and Artistic Works. (1979). http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html [accessed January 2012].

Bird, S. (2001). *White Paper on Machine-Readable Rights Information.* [cached copy]. http://web.archive.org/web/20101003110937/http://www.ldc.upenn.edu/sb/oai/rights.html

Bizer, C., Heath, T., Berners-Lee, T. (2009). *Linked Data - The Story So Far.* Available at: http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf
Brinker, S. *7 business models for linked data* and *The 8th linked data business model* (Jan2010) http://www.chiefmartec.com/2010/01/7-business-models-for-linked-data.html
David Walters, & Mark Rainbird. (2007). Cooperative innovation: a value chain approach. Journal of Enterprise Information Management, 20(5), 595-607.

Dierickx, B. and Vissers, R. (2009). Overview of IPR legislation in relation to the objectives of Europeana. [online] http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents [accessed July 2011].

EC (DG Regional Policy). (2003). *Guidelines for Successful Public-Private Partnerships.* Available at: http://ec.europa.eu/regional_policy/sources/docgener/guides/ppp_en.pdf

Eric Hellman *8 One-Way Business Models for Linked Data* (Jan 2010) http://go-to-hellman.blogspot.it/2010/01/8-one-way-business-models-for-linked.html
European Commission. (2003). Guidelines for Successful Public-Private Partnerships. Online: http://ec.europa.eu/regional_policy/sources/docgener/guides/ppp_en.pdf

European Commission. (2011). The New Renaissance: Report of the 'Comité des Sages' on bringing Europe's cultural heritage online. [online]. http://dx.doi.org/10.2759/45571 Luxemberg: Publications office of the European Union.

Europeana. (2010). *Content Development Strategy for Europeana.* Available from http://pro.europeana.eu/documents/866205/0/EV1-AF-ContentDevStrategy.pdf

Europeana. (2011). *Functional Requirements: Embedded metadata in previews.* Available from: http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsPreviewMetadata

Europeana. (2012). *Collections and Data Analysis, Strategy and Plan.* Available at: http://pro.europeana.eu/documents/866067/983522/MS11+Collections+and+Data+Analysis%2C+Strategy+and+Plan+2012-2014

Gadd, E., Oppenheim, C. and Probets, S. (2003). RoMEO Studies 5: IPR issues facing OAI Data and Service Providers. [online]. http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html [accessed January 2012].

Gadd, E., Oppenheim, C. and Probets, S. (2003). The RoMEO Project: protecting metadata in an open access environment. *Ariadne*, 36. http://www.ariadne.ac.uk/issue36/romeo/ URI: http://hdl.handle.net/2134/278

Heath, T. (2008). *Where is the business value in Linked Data?* Available at: http://tomheath.com/blog/2008/09/where-is-the-business-value-in-linked-data/
Isaac, A. (2012). *Functional and Technical Specifications: Embedded metadata in previews.* Available from: http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationPreviewMetadata

JISC. (2011). SALT – Surfacing the Academic Long Tail: A JISC Activity Data project blog. Agreeing licensing of data. http://salt11.wordpress.com/2011/07/26/agreeing-licensing-of-data/

Keller, P. (2011). *Main Aspects of the Europeana Data Exchange Agreement.* Available at: http://www.slideshare.net/paulkeller/main-aspects-of-the-europeana-data-exchange-agreement

Kreutzer, T. Validity of the Creative Commons Zero 1.0 Universal Public Domain Dedication and its usability for bibliographic metadata from the perspective of German Copyright Law. [online]. http://version1.europeana.eu/web/europeana-project/newagreement/ [accessed January 2012].

Lambert, J. (2010). Database Rights. [online]. http://niptech.wordpress.com/intellectual-property/intellectual-assets/technology-2/database-rights/ [accessed January 2012].

Laurens van Krevelen. (1987). The rights of the publisher in the information society. *Journal of Information Science* December 13: 345-351, doi:10.1177/016555158701300605

Leigh Dodds *Thoughts on Linked Data Business Models* Jan 2010)
http://www.ldodds.com/blog/2010/01/thoughts-on-linked-data-business-models/

Linked Heritage. (2011).*D4.1 Best practice report - Public Private Partnership.* Available at: http://www.linkedheritage.org/getFile.php?id=283

Linked Heritage. (2012). *D4.2 Specification of the technologies chosen.* Available at: http://www.linkedheritage.org/getFile.php?id=394

Lubart, J. L. (2011). Copyright and metadata jurisprudence. *Intellectual Property Law Newsletter.* Pennsylvania Bar Association. Spring. Available at: http://www.raderfishman.com/wp-content/uploads/2011/06/PSG_PA-Bar-Assoc_Cloudy-More-Clouds_Spring2011.pdf

Nicholas Joint. (2007). Digital information and the "privatisation of knowledge". Library Review, 56(8), 659-665.

OKFN. (2011). Principles on Open Bibliographic Data. http://openbiblio.net/files/2011/01/POBD.pdf [accessed January 2012].

Open Bibliographic Data Guide: Rights and Licensing http://obd.jisc.ac.uk/rights-and-licensing

Open Data Commons: Legal tools for Open Data. Licenses FAQ. http://opendatacommons.org/faq/licenses/ Open Knowledge Foundation.

Paskin N, 2006. *Naming and Meaning: Key to the Management of Intellectual Property in Digital Media*, in *Citeseer* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.178.6630.

Rust, G. (1998). *Metadata: the right approach. An integrated model for descriptive and rights metadata in e-commerce*. [online]. http://www.dlib.org/dlib/july98/rust/07rust.html

SABIP. (2010). *The Relationship Between Copyright and Contract Law.* Available at: http://www.ipo.gov.uk/ipresearch-relation-201007.pdf

Savenije, B. and Beunen, A. (2012). Cultural Heritage and the Public Domain. *LIBER Quarterly.* 22(2).

Vafopoulos, Michalis N., *A Framework for Linked Data Business Models* (May 2011) http://ssrn.com/abstract=1850365
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

Verwayen H, Arnoldus M, Kaufman, PB, 2011. *The Problem of the Yellow Milkmaid. A Business Model Perspective on Open Metadata*, Europeana White Paper n. 2, Nov 2011, available at http://pro.europeana.eu/web/guest/publications.

Verwayen, Arnoldus and Kaufman. (2011). *The Problem of the Yellow Milkmaid: A business model perspective on open metadata.* Second Europeana White Paper. Available at: http://pro.europeana.eu/documents/858566/2cbf1f78-e036-4088-af25-94684ff90dc5

W3C, (2011). *Library Linked Data Incubator Group Final Report.* Available at: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

Weinberger, D.. (2011, November). Letting data out of its box. *KM World*, 20(10), 1,24.

Woodward H, Needham P, 2010. *Vision for E-Book Metadata*. JISC Resource Discovery task force (RDTF) document. http://rdtf.jiscinvolve.org/wp/meeting-3/

Zepheira. (2012). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services.* Available from: http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf

11 APPENDIX – BOOKS SUBJECT CLASSIFICATION HEADINGS FOR DEFINITION OF A DATASET TO CONTRIBUTE TO EUROPEANA

11.1 BISAC

BISAC 2009 headings[136] which are directly relevant:

BISAC 2009	BISAC2009 Literal
ARC005030	ARCHITECTURE / History / Medieval
ARC005040	ARCHITECTURE / History / Renaissance
ARC005050	ARCHITECTURE / History / Baroque & Rococo
ARC005060	ARCHITECTURE / History / Romanticism
ART015030	ART / European
ART015070	ART / History / Medieval
ART015080	ART / History / Renaissance
ART015090	ART / History / Baroque & Rococo
CKB034000	COOKING / Regional & Ethnic / French
CKB036000	COOKING / Regional & Ethnic / German
CKB038000	COOKING / Regional & Ethnic / Greek
CKB043000	COOKING / Regional & Ethnic / Hungarian
CKB046000	COOKING / Regional & Ethnic / Irish
CKB047000	COOKING / Regional & Ethnic / Italian

[136] Note that any use of the BISAC codes in any database is licensable and not free; the latest version is the 2012 edition, available from: http://www.bisg.org/publications/product.php?p=14&c=437

CKB049000	COOKING / Regional & Ethnic / Jewish & Kosher
CKB055000	COOKING / Regional & Ethnic / Mediterranean
CKB065000	COOKING / Regional & Ethnic / Polish
CKB066000	COOKING / Regional & Ethnic / Portuguese
CKB074000	COOKING / Regional & Ethnic / Scandinavian
CKB080000	COOKING / Regional & Ethnic / Spanish
CKB092000	COOKING / Regional & Ethnic / European
DRA003000	DRAMA / English, Irish, Scottish, Welsh
DRA004000	DRAMA / Continental European
DRA006000	DRAMA / Ancient, Classical & Medieval
DRA010000	DRAMA / Shakespeare
DRA016000	DRAMA / Russian & Former Soviet Union*
HIS005000	HISTORY / Europe / Baltic States
HIS010000	HISTORY / Europe / General
HIS010010	HISTORY / Europe / Eastern
HIS010020	HISTORY / Europe / Western
HIS012000	HISTORY / Europe / Former Soviet Republics
HIS013000	HISTORY / Europe / France
HIS014000	HISTORY / Europe / Germany
HIS015000	HISTORY / Europe / Great Britain
HIS018000	HISTORY / Europe / Ireland

HIS020000	HISTORY / Europe / Italy
HIS022000	HISTORY / Jewish
HIS032000	HISTORY / Europe / Russia & the Former Soviet Union
HIS037010	HISTORY / Medieval
HIS037020	HISTORY / Renaissance
HIS040000	HISTORY / Europe / Austria & Hungary
HIS041010	HISTORY / Caribbean & West Indies / Cuba
HIS043000	HISTORY / Holocaust
HIS044000	HISTORY / Europe / Scandinavia
HIS045000	HISTORY / Europe / Spain & Portugal
JNF025070	JUVENILE NONFICTION / History / Europe
JNF025090	JUVENILE NONFICTION / History / Holocaust
JNF025100	JUVENILE NONFICTION / History / Medieval
JNF025160	JUVENILE NONFICTION / History / Renaissance
JNF026110	JUVENILE NONFICTION / Holidays & Celebrations / Hanukkah
JNF026120	JUVENILE NONFICTION / Holidays & Celebrations / Passover
JNF036010	JUVENILE NONFICTION / Music / Classical
JNF036020	JUVENILE NONFICTION / Music / History
JNF038060	JUVENILE NONFICTION / People & Places / Europe
JUV004010	JUVENILE FICTION / Biographical / European
JUV016040	JUVENILE FICTION / Historical / Europe

JUV016060	JUVENILE FICTION / Historical / Holocaust
JUV016070	JUVENILE FICTION / Historical / Medieval
JUV016100	JUVENILE FICTION / Historical / Renaissance
JUV017110	JUVENILE FICTION / Holidays & Celebrations / Hanukkah
JUV017120	JUVENILE FICTION / Holidays & Celebrations / Passover
JUV022020	JUVENILE FICTION / Legends, Myths, Fables / Greek & Roman
JUV022030	JUVENILE FICTION / Legends, Myths, Fables / Norse
JUV030050	JUVENILE FICTION / People & Places / Europe
LCO008000	LITERARY COLLECTIONS / Continental European
LCO009000	LITERARY COLLECTIONS / English, Irish, Scottish, Welsh
LCO014000	LITERARY COLLECTIONS / Russian & Former Soviet Union*
POE005020	POETRY / English, Irish, Scottish, Welsh
POE005030	POETRY / Continental European
POE016000	POETRY / Russian & Former Soviet Union*
REL034040	RELIGION / Holidays / Jewish

BISAC 2009 headings relevant only when combined with one or more European qualifiers:

BISAC 2009	BISAC2009 Literal
ANT000000	ANTIQUES & COLLECTIBLES / General
ANT002000	ANTIQUES & COLLECTIBLES / Art
ANT003000	ANTIQUES & COLLECTIBLES / Autographs

BISAC 2009	BISAC2009 Literal
ANT005000	ANTIQUES & COLLECTIBLES / Books
ANT006000	ANTIQUES & COLLECTIBLES / Bottles
ANT007000	ANTIQUES & COLLECTIBLES / Buttons & Pins
ANT010000	ANTIQUES & COLLECTIBLES / Clocks & Watches
ANT011000	ANTIQUES & COLLECTIBLES / Coins, Currency & Medals
ANT012000	ANTIQUES & COLLECTIBLES / Comics
ANT015000	ANTIQUES & COLLECTIBLES / Dolls
ANT016000	ANTIQUES & COLLECTIBLES / Firearms & Weapons
ANT017000	ANTIQUES & COLLECTIBLES / Furniture
ANT018000	ANTIQUES & COLLECTIBLES / Glass & Glassware
ANT021000	ANTIQUES & COLLECTIBLES / Jewelry
ANT022000	ANTIQUES & COLLECTIBLES / Kitchenware
ANT023000	ANTIQUES & COLLECTIBLES / Magazines & Newspapers
ANT024000	ANTIQUES & COLLECTIBLES / Military
ANT025000	ANTIQUES & COLLECTIBLES / Performing Arts
ANT028000	ANTIQUES & COLLECTIBLES / Non-Sports Cards
ANT029000	ANTIQUES & COLLECTIBLES / Paper Ephemera
ANT031000	ANTIQUES & COLLECTIBLES / Political
ANT032000	ANTIQUES & COLLECTIBLES / Porcelain & China
ANT033000	ANTIQUES & COLLECTIBLES / Postcards

BISAC 2009	BISAC2009 Literal
ANT034000	ANTIQUES & COLLECTIBLES / Posters
ANT035000	ANTIQUES & COLLECTIBLES / Pottery & Ceramics
ANT036000	ANTIQUES & COLLECTIBLES / Radios & Televisions (see also Performing Arts)
ANT037000	ANTIQUES & COLLECTIBLES / Records
ANT040000	ANTIQUES & COLLECTIBLES / Rugs
ANT041000	ANTIQUES & COLLECTIBLES / Silver, Gold & Other Metals
ANT042000	ANTIQUES & COLLECTIBLES / Sports Cards / General
ANT042010	ANTIQUES & COLLECTIBLES / Sports Cards / Baseball
ANT042020	ANTIQUES & COLLECTIBLES / Sports Cards / Basketball
ANT042030	ANTIQUES & COLLECTIBLES / Sports Cards / Football
ANT042040	ANTIQUES & COLLECTIBLES / Sports Cards / Hockey
ANT043000	ANTIQUES & COLLECTIBLES / Sports (see also headings under Sports Cards)
ANT044000	ANTIQUES & COLLECTIBLES / Stamps
ANT045000	ANTIQUES & COLLECTIBLES / Teddy Bears
ANT047000	ANTIQUES & COLLECTIBLES / Textiles & Costume
ANT049000	ANTIQUES & COLLECTIBLES / Toy Animals
ANT050000	ANTIQUES & COLLECTIBLES / Toys
ANT052000	ANTIQUES & COLLECTIBLES / Popular Culture
ANT053000	ANTIQUES & COLLECTIBLES / Figurines
ANT055000	ANTIQUES & COLLECTIBLES / Tobacco-Related

BISAC 2009	BISAC2009 Literal
ARC000000	ARCHITECTURE / General
ARC001000	ARCHITECTURE / Criticism
ARC002000	ARCHITECTURE / Decoration & Ornament
ARC003000	ARCHITECTURE / Buildings / Residential
ARC005000	ARCHITECTURE / History / General
ARC005010	ARCHITECTURE / History / Prehistoric & Primitive
ARC005020	ARCHITECTURE / History / Ancient & Classical
ARC005030	ARCHITECTURE / History / Medieval
ARC005040	ARCHITECTURE / History / Renaissance
ARC005050	ARCHITECTURE / History / Baroque & Rococo
ARC005060	ARCHITECTURE / History / Romanticism
ARC005070	ARCHITECTURE / History / Modern (late 19th Century to 1945)
ARC005080	ARCHITECTURE / History / Contemporary (1945-)
ARC007000	ARCHITECTURE / Interior Design / General
ARC007010	ARCHITECTURE / Interior Design / Lighting
ARC008000	ARCHITECTURE / Landscape
ARC016000	ARCHITECTURE / Buildings / Religious
ARC020000	ARCHITECTURE / Regional
ARC024000	ARCHITECTURE / Buildings / General
ARC024010	ARCHITECTURE / Buildings / Landmarks & Monuments

BISAC 2009	BISAC2009 Literal
ART000000	ART / General
ART006000	ART / Collections, Catalogs, Exhibitions / General
ART006010	ART / Collections, Catalogs, Exhibitions / Group Shows
ART006020	ART / Collections, Catalogs, Exhibitions / Permanent Collections
ART013000	ART / Folk & Outsider Art
ART015000	ART / History / General
ART015050	ART / History / Prehistoric & Primitive
ART015060	ART / History / Ancient & Classical
ART015070	ART / History / Medieval
ART015080	ART / History / Renaissance
ART015090	ART / History / Baroque & Rococo
ART015100	ART / History / Modern (late 19th Century to 1945)
ART015110	ART / History / Contemporary (1945-)
ART015120	ART / History / Romanticism
ART016000	ART / Individual Artists / General
ART016010	ART / Individual Artists / Artists' Books
ART016020	ART / Individual Artists / Essays
ART016030	ART / Individual Artists / Monographs
ART026000	ART / Sculpture & Installation
ART035000	ART / Subjects & Themes / Religious

BISAC 2009	BISAC2009 Literal
ART037000	ART / Art & Politics
ART045000	ART / Ceramics
ART046000	ART / Digital
ART048000	ART / Prints
ART049000	ART / Russian & Former Soviet Union
ART050000	ART / Subjects & Themes / General
ART050010	ART / Subjects & Themes / Human Figure
ART050020	ART / Subjects & Themes / Landscapes
ART050030	ART / Subjects & Themes / Plants & Animals
ART050040	ART / Subjects & Themes / Portraits
ART054000	ART / Annuals
ART055000	ART / Body Art & Tattooing
ART057000	ART / Film & Video
ART058000	ART / Graffiti & Street Art
ART060000	ART / Performance
BIO000000	BIOGRAPHY & AUTOBIOGRAPHY / General
BIO001000	BIOGRAPHY & AUTOBIOGRAPHY / Artists, Architects, Photographers
BIO002000	BIOGRAPHY & AUTOBIOGRAPHY / Cultural Heritage
BIO003000	BIOGRAPHY & AUTOBIOGRAPHY / Business
BIO004000	BIOGRAPHY & AUTOBIOGRAPHY / Composers & Musicians

BISAC 2009	BISAC2009 Literal
BIO005000	BIOGRAPHY & AUTOBIOGRAPHY / Entertainment & Performing Arts
BIO006000	BIOGRAPHY & AUTOBIOGRAPHY / Historical
BIO007000	BIOGRAPHY & AUTOBIOGRAPHY / Literary
BIO008000	BIOGRAPHY & AUTOBIOGRAPHY / Military
BIO009000	BIOGRAPHY & AUTOBIOGRAPHY / Philosophers
BIO010000	BIOGRAPHY & AUTOBIOGRAPHY / Political
BIO011000	BIOGRAPHY & AUTOBIOGRAPHY / Presidents & Heads of State
BIO013000	BIOGRAPHY & AUTOBIOGRAPHY / Rich & Famous
BIO014000	BIOGRAPHY & AUTOBIOGRAPHY / Royalty
BIO015000	BIOGRAPHY & AUTOBIOGRAPHY / Science & Technology
BIO016000	BIOGRAPHY & AUTOBIOGRAPHY / Sports
BIO017000	BIOGRAPHY & AUTOBIOGRAPHY / Medical
BIO018000	BIOGRAPHY & AUTOBIOGRAPHY / Religious
BIO019000	BIOGRAPHY & AUTOBIOGRAPHY / Educators
BIO020000	BIOGRAPHY & AUTOBIOGRAPHY / Lawyers & Judges
BIO021000	BIOGRAPHY & AUTOBIOGRAPHY / Social Scientists & Psychologists
BIO022000	BIOGRAPHY & AUTOBIOGRAPHY / Women
BIO023000	BIOGRAPHY & AUTOBIOGRAPHY / Adventurers & Explorers
BIO024000	BIOGRAPHY & AUTOBIOGRAPHY / Criminals & Outlaws
BIO025000	BIOGRAPHY & AUTOBIOGRAPHY / Editors, Journalists, Publisher

BISAC 2009	BISAC2009 Literal
BIO026000	BIOGRAPHY & AUTOBIOGRAPHY / Personal Memoirs
BIO027000	BIOGRAPHY & AUTOBIOGRAPHY / Law Enforcement
BIO028000	BIOGRAPHY & AUTOBIOGRAPHY / Native Americans
BUS023000	BUSINESS & ECONOMICS / Economic History
BUS077000	BUSINESS & ECONOMICS / Corporate & Business History
CGN000000	COMICS & GRAPHIC NOVELS / General
CGN001000	COMICS & GRAPHIC NOVELS / Anthologies
CGN006000	COMICS & GRAPHIC NOVELS / Literary
CGN007000	COMICS & GRAPHIC NOVELS / Nonfiction
CGN008000	COMICS & GRAPHIC NOVELS / Contemporary Women
CKB011000	COOKING / Regional & Ethnic / English, Scottish & Welsh
CKB030000	COOKING / Essays
CKB041000	COOKING / History
CKB042000	COOKING / Holiday
CKB045000	COOKING / Regional & Ethnic / International
CKB100000	COOKING / Beverages / General
CKB101000	COOKING / Courses & Dishes / General
COM079000	COMPUTERS / Social Aspects / General
COM080000	COMPUTERS / History
CRA000000	CRAFTS & HOBBIES / General

BISAC 2009	BISAC2009 Literal
DES008000	DESIGN / History & Criticism
DRA000000	DRAMA / General
DRA002000	DRAMA / Anthologies (multiple authors)
DRA008000	DRAMA / Religious & Liturgical
EDU000000	EDUCATION / General
EDU016000	EDUCATION / History
EDU034000	EDUCATION / Educational Policy & Reform / General
EDU034010	EDUCATION / Educational Policy & Reform / School Safety & Violence
FAM000000	FAMILY & RELATIONSHIPS / General
FIC004000	FICTION / Classics
FIC009030	FICTION / Fantasy / Historical
FIC027050	FICTION / Romance / Historical
FIC042010	FICTION / Christian / Classic & Allegory
FIC042030	FICTION / Christian / Historical
GAM000000	GAMES / General
GAR000000	GARDENING / General
GAR019000	GARDENING / Regional / General
HIS000000	HISTORY / General
HIS022000	HISTORY / Jewish
HIS027000	HISTORY / Military / General

BISAC 2009	BISAC2009 Literal
HIS037030	HISTORY / Modern / General
HIS037040	HISTORY / Modern / 17th Century
HIS037050	HISTORY / Modern / 18th Century
HIS037060	HISTORY / Modern / 19th Century
HIS037070	HISTORY / Modern / 20th Century
HIS037080	HISTORY / Modern / 21st Century
HIS037090	HISTORY / Modern / 16th Century
HIS039000	HISTORY / Civilization
HIS049000	HISTORY / Essays
HIS051000	HISTORY / Expeditions & Discoveries
HIS052000	HISTORY / Historical Geography
HIS054000	HISTORY / Social History
HOM000000	HOUSE & HOME / General
HUM000000	HUMOR / General
JNF000000	JUVENILE NONFICTION / General
JNF004000	JUVENILE NONFICTION / Antiques & Collectibles
JNF005000	JUVENILE NONFICTION / Architecture
JNF006000	JUVENILE NONFICTION / Art / General
JNF006040	JUVENILE NONFICTION / Art / History
JNF007010	JUVENILE NONFICTION / Biography & Autobiography / Art

BISAC 2009	BISAC2009 Literal
JNF007020	JUVENILE NONFICTION / Biography & Autobiography / Historical
JNF007030	JUVENILE NONFICTION / Biography & Autobiography / Literary
JNF007040	JUVENILE NONFICTION / Biography & Autobiography / Music
JNF007050	JUVENILE NONFICTION / Biography & Autobiography / Cultural Heritage
JNF007060	JUVENILE NONFICTION / Biography & Autobiography / Performing Arts
JNF007070	JUVENILE NONFICTION / Biography & Autobiography / Political
JNF007080	JUVENILE NONFICTION / Biography & Autobiography / Religious (see also Religious / Christian / Biography & Autobiography)
JNF007110	JUVENILE NONFICTION / Biography & Autobiography / Social Activists
JNF007120	JUVENILE NONFICTION / Biography & Autobiography / Women
JNF025140	JUVENILE NONFICTION / History / Modern
JNF025160	JUVENILE NONFICTION / History / Renaissance
JNF026000	JUVENILE NONFICTION / Holidays & Celebrations / General (see also Religious / Christian / Holidays & Celebrations)
JNF026010	JUVENILE NONFICTION / Holidays & Celebrations / Christmas & Advent
JNF026020	JUVENILE NONFICTION / Holidays & Celebrations / Easter & Lent
JNF026030	JUVENILE NONFICTION / Holidays & Celebrations / Halloween
JNF026070	JUVENILE NONFICTION / Holidays & Celebrations / Valentine's Day
JNF026090	JUVENILE NONFICTION / Holidays & Celebrations / Other, Religious
JNF026130	JUVENILE NONFICTION / Holidays & Celebrations / Patriotic Holidays
JNF028000	JUVENILE NONFICTION / Humor / General
JNF031000	JUVENILE NONFICTION / Lifestyles / City & Town Life

BISAC 2009	BISAC2009 Literal
JNF032000	JUVENILE NONFICTION / Lifestyles / Country Life
JNF033000	JUVENILE NONFICTION / Lifestyles / Farm & Ranch Life
JNF036000	JUVENILE NONFICTION / Music / General
JNF036040	JUVENILE NONFICTION / Music / Jazz
JNF036050	JUVENILE NONFICTION / Music / Popular
JNF036060	JUVENILE NONFICTION / Music / Rap & Hip Hop
JNF036070	JUVENILE NONFICTION / Music / Rock
JNF039000	JUVENILE NONFICTION / Performing Arts / General
JNF039010	JUVENILE NONFICTION / Performing Arts / Circus
JNF039020	JUVENILE NONFICTION / Performing Arts / Dance
JNF039030	JUVENILE NONFICTION / Performing Arts / Film
JNF039040	JUVENILE NONFICTION / Performing Arts / Television & Radio
JNF039050	JUVENILE NONFICTION / Performing Arts / Theater
JNF040000	JUVENILE NONFICTION / Philosophy
JNF041000	JUVENILE NONFICTION / Photography
JNF042000	JUVENILE NONFICTION / Poetry / General
JNF049080	JUVENILE NONFICTION / Religion / Christianity
JNF049100	JUVENILE NONFICTION / Religion / Islam
JNF049110	JUVENILE NONFICTION / Religion / Judaism
JNF049130	JUVENILE NONFICTION / Religious / Christian / General

BISAC 2009	BISAC2009 Literal
JNF049240	JUVENILE NONFICTION / Religious / Christian / Holidays & Celebrations
JNF049270	JUVENILE NONFICTION / Religious / Christian / People & Places
JNF049290	JUVENILE NONFICTION / Religious / Christian / Social Issues
JNF052000	JUVENILE NONFICTION / Social Science / General
JNF052010	JUVENILE NONFICTION / Social Science / Archaeology
JNF052020	JUVENILE NONFICTION / Social Science / Customs, Traditions, Anthropology
JNF052030	JUVENILE NONFICTION / Social Science / Folklore & Mythology
JNF053000	JUVENILE NONFICTION / Social Issues / General (see also headings under Family)
JNF054000	JUVENILE NONFICTION / Sports & Recreation / General
JNF059000	JUVENILE NONFICTION / Clothing & Dress
JNF062000	JUVENILE NONFICTION / Comics & Graphic Novels / General
JNF062010	JUVENILE NONFICTION / Comics & Graphic Novels / Biography
JNF062020	JUVENILE NONFICTION / Comics & Graphic Novels / History
JUV002270	JUVENILE FICTION / Animals / Mythical
JUV003000	JUVENILE FICTION / Art & Architecture
JUV004000	JUVENILE FICTION / Biographical / General
JUV007000	JUVENILE FICTION / Classics
JUV011050	JUVENILE FICTION / People & Places / United States / Other
JUV012000	JUVENILE FICTION / Fairy Tales & Folklore / Anthologies
JUV012020	JUVENILE FICTION / Fairy Tales & Folklore / Country & Ethnic

BISAC 2009	BISAC2009 Literal
JUV012030	JUVENILE FICTION / Fairy Tales & Folklore / General
JUV012040	JUVENILE FICTION / Fairy Tales & Folklore / Adaptations
JUV016000	JUVENILE FICTION / Historical / General
JUV016020	JUVENILE FICTION / Historical / Ancient Civilizations
JUV016050	JUVENILE FICTION / Historical / Exploration & Discovery
JUV016080	JUVENILE FICTION / Historical / Military & Wars
JUV016130	JUVENILE FICTION / Historical / Other
JUV017000	JUVENILE FICTION / Holidays & Celebrations / General (see also Religious / Christian / Holidays & Celebrations)
JUV017010	JUVENILE FICTION / Holidays & Celebrations / Christmas & Advent
JUV017020	JUVENILE FICTION / Holidays & Celebrations / Easter & Lent
JUV017030	JUVENILE FICTION / Holidays & Celebrations / Halloween
JUV017070	JUVENILE FICTION / Holidays & Celebrations / Valentine's Day
JUV017080	JUVENILE FICTION / Holidays & Celebrations / Other, Non-Religious
JUV017090	JUVENILE FICTION / Holidays & Celebrations / Other, Religious
JUV017130	JUVENILE FICTION / Holidays & Celebrations / Patriotic Holidays
JUV022000	JUVENILE FICTION / Legends, Myths, Fables / General
JUV022010	JUVENILE FICTION / Legends, Myths, Fables / Arthurian
JUV022040	JUVENILE FICTION / Legends, Myths, Fables / Other
JUV023000	JUVENILE FICTION / Lifestyles / City & Town Life
JUV024000	JUVENILE FICTION / Lifestyles / Country Life

BISAC 2009	BISAC2009 Literal
JUV025000	JUVENILE FICTION / Lifestyles / Farm & Ranch Life
JUV030000	JUVENILE FICTION / People & Places / General
LCO000000	LITERARY COLLECTIONS / General
LCO003000	LITERARY COLLECTIONS / Ancient, Classical & Medieval
LCO010000	LITERARY COLLECTIONS / Essays
LCO011000	LITERARY COLLECTIONS / Letters
MED039000	MEDICAL / History
MUS000000	MUSIC / General
MUS002000	MUSIC / Genres & Styles / Ballet
MUS003000	MUSIC / Genres & Styles / Blues
MUS005000	MUSIC / Genres & Styles / Chamber
MUS006000	MUSIC / Genres & Styles / Classical
MUS009000	MUSIC / Religious / Contemporary Christian
MUS010000	MUSIC / Genres & Styles / Country & Bluegrass
MUS011000	MUSIC / Genres & Styles / Dance
MUS013000	MUSIC / Genres & Styles / Electronic
MUS014000	MUSIC / Ethnic
MUS017000	MUSIC / Genres & Styles / Folk & Traditional
MUS018000	MUSIC / Religious / Gospel
MUS019000	MUSIC / Genres & Styles / Heavy Metal

BISAC 2009	BISAC2009 Literal
MUS020000	MUSIC / History & Criticism
MUS021000	MUSIC / Religious / Hymns
MUS024000	MUSIC / Genres & Styles / International
MUS025000	MUSIC / Genres & Styles / Jazz
MUS026000	MUSIC / Genres & Styles / Children's
MUS027000	MUSIC / Genres & Styles / New Age
MUS028000	MUSIC / Genres & Styles / Opera
MUS029000	MUSIC / Genres & Styles / Pop Vocal
MUS030000	MUSIC / Genres & Styles / Punk
MUS031000	MUSIC / Genres & Styles / Rap & Hip Hop
MUS035000	MUSIC / Genres & Styles / Rock
MUS036000	MUSIC / Genres & Styles / Latin
MUS039000	MUSIC / Genres & Styles / Soul & R 'n' B
MUS045000	MUSIC / Genres & Styles / Military & Marches
MUS046000	MUSIC / Genres & Styles / Musicals
MUS047000	MUSIC / Genres & Styles / Reggae
MUS048000	MUSIC / Religious / General
MUS048010	MUSIC / Religious / Christian
MUS048020	MUSIC / Religious / Jewish
MUS048030	MUSIC / Religious / Muslim

BISAC 2009	BISAC2009 Literal
MUS049000	MUSIC / Genres & Styles / General
MUS050000	MUSIC / Individual Composer & Musician
MUS051000	MUSIC / Genres & Styles / Choral
PER000000	PERFORMING ARTS / General
PER003000	PERFORMING ARTS / Dance / General
PER004000	PERFORMING ARTS / Film & Video / General
PER004030	PERFORMING ARTS / Film & Video / History & Criticism
PER008000	PERFORMING ARTS / Radio / General
PER008010	PERFORMING ARTS / Radio / History & Criticism
PER010000	PERFORMING ARTS / Television / General
PER010030	PERFORMING ARTS / Television / History & Criticism
PER011000	PERFORMING ARTS / Theater / General
PER011020	PERFORMING ARTS / Theater / History & Criticism
PER013000	PERFORMING ARTS / Theater / Broadway & Musical Revue
PER015000	PERFORMING ARTS / Comedy
PER017000	PERFORMING ARTS / Animation
PER018000	PERFORMING ARTS / Individual Director (see also BIOGRAPHY & AUTOBIOGRAPHY / Entertainment & Performing Arts)
PER019000	PERFORMING ARTS / Storytelling
PHI000000	PHILOSOPHY / General
PHI002000	PHILOSOPHY / History & Surveys / Ancient & Classical

BISAC 2009	BISAC2009 Literal
PHI012000	PHILOSOPHY / History & Surveys / Medieval
PHI016000	PHILOSOPHY / History & Surveys / Modern
PHI019000	PHILOSOPHY / Political
PHI035000	PHILOSOPHY / Essays*
PHO000000	PHOTOGRAPHY / General
PHO001000	PHOTOGRAPHY / Subjects & Themes / Architectural & Industrial
PHO004000	PHOTOGRAPHY / Collections, Catalogs, Exhibitions / General
PHO004010	PHOTOGRAPHY / Collections, Catalogs, Exhibitions / Group Shows
PHO004020	PHOTOGRAPHY / Collections, Catalogs, Exhibitions / Permanent Collections
PHO009000	PHOTOGRAPHY / Subjects & Themes / Fashion
PHO010000	PHOTOGRAPHY / History
PHO011000	PHOTOGRAPHY / Individual Photographers / General
PHO011010	PHOTOGRAPHY / Individual Photographers / Artists' Books
PHO011020	PHOTOGRAPHY / Individual Photographers / Essays
PHO011030	PHOTOGRAPHY / Individual Photographers / Monographs
PHO013000	PHOTOGRAPHY / Subjects & Themes / Plants & Animals
PHO014000	PHOTOGRAPHY / Photoessays & Documentaries
PHO016000	PHOTOGRAPHY / Subjects & Themes / Portraits
PHO019000	PHOTOGRAPHY / Subjects & Themes / Regional (see also TRAVEL / Pictorials)
PHO023000	PHOTOGRAPHY / Subjects & Themes / General

BISAC 2009	BISAC2009 Literal
PHO023010	PHOTOGRAPHY / Subjects & Themes / Aerial
PHO023040	PHOTOGRAPHY / Subjects & Themes / Landscapes
PHO023060	PHOTOGRAPHY / Subjects & Themes / Sports
PHO023070	PHOTOGRAPHY / Subjects & Themes / Celebrations & Events
PHO023090	PHOTOGRAPHY / Subjects & Themes / Lifestyles
PHO023100	PHOTOGRAPHY / Subjects & Themes / Historical*
PHO024000	PHOTOGRAPHY / Techniques / Digital (see also COMPUTERS / Digital Media / Photography)
POE000000	POETRY / General
POE001000	POETRY / Anthologies (multiple authors)
POE003000	POETRY / Inspirational & Religious
POE008000	POETRY / Ancient, Classical & Medieval
POE014000	POETRY / Epic
POL000000	POLITICAL SCIENCE / General
POL010000	POLITICAL SCIENCE / History & Theory
POL032000	POLITICAL SCIENCE / Essays
PSY015000	PSYCHOLOGY / History
REL006630	RELIGION / Biblical Studies / History & Culture
REL007010	RELIGION / Buddhism / History
REL013000	RELIGION / Christianity / Literature & the Arts
REL015000	RELIGION / Christianity / History

BISAC 2009	BISAC2009 Literal
REL032010	RELIGION / Hinduism / History
REL033000	RELIGION / History
REL034000	RELIGION / Holidays / General
REL034010	RELIGION / Holidays / Christian
REL034020	RELIGION / Holidays / Christmas & Advent
REL034030	RELIGION / Holidays / Easter & Lent
REL034050	RELIGION / Holidays / Other
REL037010	RELIGION / Islam / History
REL040030	RELIGION / Judaism / History
REL067080	RELIGION / Christian Theology / History
REL070000	RELIGION / Christianity / General
REL072000	RELIGION / Antiquities & Archaeology
REL084000	RELIGION / Religion, Politics & State
REL108020	RELIGION / Christian Church / History
REL110000	RELIGION / Christianity / Saints & Sainthood
SCI034000	SCIENCE / History
SOC000000	SOCIAL SCIENCE / General
SOC002010	SOCIAL SCIENCE / Anthropology / Cultural
SOC003000	SOCIAL SCIENCE / Archaeology
SOC005000	SOCIAL SCIENCE / Customs & Traditions

BISAC 2009	BISAC2009 Literal
SOC006000	SOCIAL SCIENCE / Demography
SOC007000	SOCIAL SCIENCE / Emigration & Immigration
SOC011000	SOCIAL SCIENCE / Folklore & Mythology
SOC013000	SOCIAL SCIENCE / Gerontology
SOC014000	SOCIAL SCIENCE / Holidays (non-religious)
SOC020000	SOCIAL SCIENCE / Minority Studies
SOC022000	SOCIAL SCIENCE / Popular Culture
SOC026000	SOCIAL SCIENCE / Sociology / General
SOC026010	SOCIAL SCIENCE / Sociology / Marriage & Family
SOC026020	SOCIAL SCIENCE / Sociology / Rural
SOC026030	SOCIAL SCIENCE / Sociology / Urban
SOC041000	SOCIAL SCIENCE / Essays
SPO000000	SPORTS & RECREATION / General
TRA001050	TRANSPORTATION / Automotive / History
TRA002010	TRANSPORTATION / Aviation / History
TRA003010	TRANSPORTATION / Motorcycles / History
TRA004010	TRANSPORTATION / Railroads / History
TRA006010	TRANSPORTATION / Ships & Shipbuilding / History
TRV016000	TRAVEL / Museums, Tours, Points of Interest
TRV026060	TRAVEL / Special Interest / Religious

BISAC 2009	BISAC2009 Literal
TRV026090	TRAVEL / Special Interest / Literary*

11.2 BIC

Directly relevant BIC version 2.1 headings:

Code	Heading	V2.1
ACG	History of art: ancient & classical art, BCE to c 500 CE	
ACK	History of art: Byzantine & Medieval art c 500 CE to c 1400	
ACND	Renaissance art	
ACNH	Art & design styles: Mannerism	
ACQB	Art & design styles: Baroque	
ACQH	Art & design styles: Classicism	
ACVC	Art & design styles: Romanticism	
ACVM	Art & design styles: Pre-Raphaelite art	
ACVN	Art & design styles: Arts & Crafts style	
ACVT	Art & design styles: Impressionism & Post-Impressionism	
ACVY	Art & design styles: Art Nouveau	
ACXD1	Art & design styles: Expressionism	
ACXD3	Art & design styles: Art Deco	
ACXD5	Art & design styles: Cubism	

AFTC	Celtic arts & crafts	New
AVGC	Western "classical" music	
AVGC1	Early music (up to c 1000 CE)	
AVGC2	Medieval & Renaissance music (c 1000 to c 1600)	
AVGC3	Baroque music (c 1600 to c 1750)	
AVGC4	Classical music (c 1750 to c 1830)	
DDS	Shakespeare plays	
DSBB	Literary studies: classical, early & medieval	
DSBD	Literary studies: c 1500 to c 1800	
DSGS	Shakespeare studies & criticism	
HBJD	European history	
HBJD1	British & Irish history	
HBJD2	Storia d'Italia	CCE only
HBLC1	Medieval history	
HBTV2	French Revolution	
HBTV4	Russian Revolution	
HBTW	The Cold War	
HBTZ1	The Holocaust	
HBWC	Crusades	
HBWE	English Civil War	
HBWH	Napoleonic Wars	

HBWL	Crimean War	
HBWN	First World War	
HBWP	Spanish Civil War	
HBWQ	Second World War	
HDDK	Classical Greek & Roman archaeology	
HDDM	Medieval European archaeology	
HPCA	Western philosophy: Ancient, to c 500	
HPCB	Western philosophy: Medieval & Renaissance, c 500 to c 1600	
HRKP2	Celtic religion & mythology	
HRKP3	Ancient Greek religion & mythology	New
HRKP4	Roman religion & mythology	New
HRKP5	Norse religion & mythology	New
JFSR1	Jewish studies	

BIC headings relevant when combined with one or more relevant qualifiers:

Code	Heading
A	The arts
AB	The arts: general issues
ABK	Forgery, falsification & theft of artworks
AC	History of art / art & design styles

Code	Heading
ACN	History of art & design styles: c 1400 to c 1600
ACQ	History of art & design styles: c 1600 to c 1800
ACV	History of art & design styles: c 1800 to c 1900
ACVC	Art & design styles: Romanticism
ACX	History of art & design styles: from c 1900 -
ACXD	Art & design styles: c 1900 to c 1960
ACXD2	Art & design styles: Modernist design & Bauhaus
ACXD7	Art & design styles: Surrealism & Dada
ACXD9	Art & design styles: Abstract Expressionism
ACXJ	Art & design styles: from c 1960
ACXJ1	Art & design styles: Pop art
ACXJ4	Art & design styles: Minimalism
ACXJ5	Art & design styles: Conceptual art
ACXJ8	Art & design styles: Postmodernism
AF	Art forms
AFC	Painting & paintings
AFCC	Watercolours
AFCL	Oils
AFF	Drawing & drawings
AFH	Prints & printmaking

Code	Heading
AFJ	Other graphic art forms
AFJD	Collage & photomontage
AFJG	Graffiti & street art
AFK	Non-graphic art forms
AFKB	Sculpture
AFKC	Carvings: artworks
AFKG	Precious metal, precious stones & jewellery: artworks & design
AFKN	Installation art
AFKP	Performance art
AFKV	Electronic, holographic & video art
AFP	Ceramic arts, pottery, glass
AFPC	Ceramics: artworks
AFPM	Mosaics: artworks
AFPS	Stained glass: artworks
AFT	Decorative arts
AFTB	Folk art
AFW	Textile artworks
AFWD	Textile artworks: carpets & rugs
AFWH	Textile artworks: tapestries, hangings & quilts
AFY	Body art & tattooing

Code	Heading
AG	Art treatments & subjects
AGB	Individual artists, art monographs
AGC	Exhibition catalogues & specific collections
AGH	Human figures depicted in art
AGHF	Portraits in art
AGHN	Nudes depicted in art
AGK	Small-scale, secular & domestic scenes in art
AGN	Animals & nature in art (still life, landscapes & seascapes, etc)
AGNB	Botanical art
AGP	Man-made objects depicted in art (cityscapes, machines, etc)
AGR	Religious subjects depicted in art
AGZ	Art techniques & principles
AJ	Photography & photographs
AJB	Individual photographers
AJC	Photographs: collections
AJCP	Photographs: portraits
AJCR	Photographic reportage
AJR	Special kinds of photography
AJRD	Cinematography, television camerawork
AJRH	Video photography

Code	Heading
AJRK	Aerial photography
AK	Industrial / commercial art & design
AKB	Individual designers
AKC	Graphic design
AKD	Typography & lettering
AKH	Book design
AKL	Illustration & commercial art
AKLB	Illustration
AKLC	Comic book & cartoon art
AKLC1	Graphic novel & Manga artwork
AKLP	Poster art
AKP	Product design
AKR	Furniture design
AKT	Fashion & textiles: design
AKTH	History of fashion
AKTX	Textile design & theory
AM	Architecture
AMB	Individual architects & architectural firms
AMG	Public buildings: civic, commercial, industrial, etc
AMGC	Concert halls, arenas, stadia

Code	Heading
AMGD	Memorials, monuments
AMK	Residential buildings, domestic buildings
AMKD	Houses, apartments, flats, etc
AMKH	Palaces, chateaux, country houses
AMKL	Castles & fortifications
AMN	Religious buildings
AMR	Professional interior design
AMV	Landscape art & architecture
AMVD	City & town planning - architectural aspects
AMX	History of architecture
AN	Theatre studies
ANB	Theatre: individual actors & directors
AP	Film, TV & radio
APB	Individual actors & performers
APF	Films, cinema
APFB	Individual film directors, film-makers
APFD	Film scripts & screenplays
APFN	Film: styles & genres
APFR	Documentary films
APFV	Animated films

Code	Heading
APT	Television
APTD	Television scripts & screenplays
APTS	Television soap operas
APW	Radio
APWD	Radio scripts
AS	Dance & other performing arts
ASD	Dance
ASDC	Choreography
ASDL	Ballet
ASDR	Ballroom dancing
ASDT	Contemporary dance
ASDX	Folk dancing
ASZ	Other performing arts
ASZB	Performing arts: comedy
ASZC	Mime
ASZD	Street theatre
ASZG	Conjuring & magic
ASZH	Variety shows, music hall, cabaret
ASZJ	Juggling
ASZM	Puppetry, miniature & toy theatre

Code	Heading
ASZP	Pageants, parades, festivals
ASZW	Circus
ASZX	Animal spectacles
AV	Music
AVG	Music: styles & genres
AVGC5	Romantic music (c 1830 to c 1900)
AVGC6	20th century & contemporary classical music
AVGC8	Choral music
AVGC9	Opera
AVGD	Sacred & religious music
AVGF	Light orchestral & big band music
AVGG	Brass band, military music & marches
AVGH	Folk & traditional music
AVGJ	Jazz
AVGK	Blues
AVGL	Country & Western music
AVGM	Musicals
AVGN	Popular music, easy listening
AVGP	Rock & Pop music
AVGQ	Soul & R 'n' B

Code	Heading
AVGR	Rap & Hip-Hop
AVGS	Reggae
AVGT	Heavy Metal music
AVGU	Punk, New Wave & Indie
AVGV	Electronic music
AVGW	World music
AVGZ	Ambient & New Age music
AVH	Individual composers & musicians, specific bands & groups
AVQ	Musical scores, lyrics & libretti
AVQS	Songbooks
AVR	Musical instruments & instrumental ensembles
AVRB	Orchestras
AVRD	Chamber ensembles
AVRG	Keyboard instruments
AVRJ	Percussion instruments
AVRL	String instruments
AVRL1	Guitar
AVRN	Wind instruments
AVRQ	Mechanical musical instruments
AVRS	Electronic musical instruments

Code	Heading
B	Biography & True Stories
BG	Biography: general
BGA	Autobiography: general
BGF	Biography: arts & entertainment
BGFA	Autobiography: arts & entertainment
BGH	Biography: historical, political & military
BGHA	Autobiography: historical, political & military
BGL	Biography: literary
BGLA	Autobiography: literary
BGR	Biography: royalty
BGRA	Autobiography: royalty
BGS	Biography: sport
BGSA	Autobiography: sport
BGT	Biography: science, technology & medicine
BGTA	Autobiography: science, technology & medicine
BGX	Biography: religious & spiritual
BGXA	Autobiography: religious & spiritual
BJ	Diaries, letters & journals
BK	Collected biographies
BM	Memoirs

Code	Heading
BT	True stories
BTH	True stories: discovery / historical / scientific
BTM	True war & combat stories
CBX	Language: history & general works
CFB	Sociolinguistics
CFFD	Dialect, slang & jargon
DB	Classical texts
DC	Poetry
DCF	Poetry by individual poets
DD	Plays, playscripts
DN	Prose: non-fiction
DNF	Literary essays
DNJ	Reportage & collected journalism
DNS	Speeches
DQ	Anthologies (non-poetry)
DS	Literature: history & criticism
DSBF	Literary studies: c 1800 to c 1900
DSBH	Literary studies: from c 1900 -
DSC	Literary studies: poetry & poets
DSG	Literary studies: plays & playwrights

Code	Heading
DSK	Literary studies: fiction, novelists & prose writers
DSR	Literary reference works
DSRC	Literary companions, book reviews & guides
DSY	Children's & teenage literature studies
FC	Classic fiction (pre c 1945)
FXZ	Graphic novels: true stories & non-fiction
FZG	Graphic novels: history & criticism
GTB	Regional studies
HB	History
HBJ	Regional & national history
HBL	History: earliest times to present day
HBLA	Ancient history: to c 500 CE
HBLA1	Classical history / classical civilisation
HBLC	Early history: c 500 to c 1450/1500
HBLH	Early modern history: c 1450/1500 to c 1700
HBLL	Modern history to 20th century: c 1700 to c 1900
HBLW	20th century history: c 1900 to c 2000
HBLW3	Postwar 20th century history, from c 1945 to c 2000
HBLX	21st century history: from c 2000 -
HBT	History: specific events & topics

Code	Heading
HBTB	Social & cultural history
HBTD	Oral history
HBTG	Genealogy, heraldry, names & honours
HBTK	Industrialisation & industrial history
HBTM	Maritime history
HBTP	Historical geography
HBTP1	Historical maps & atlases
HBTQ	Colonialism & imperialism
HBTR	National liberation & independence, post-colonialism
HBTS	Slavery & abolition of slavery
HBTV	Revolutions, uprisings, rebellions
HBTZ	Genocide & ethnic cleansing
HBW	Military history
HBWF	American War of Independence
HBWJ	American Civil War
HBWM	Boer Wars
HBWS	Military history: post WW2 conflicts
HBWS1	Korean War
HBWS2	Vietnam War
HBWS3	Gulf War

Code	Heading
HBWS4	Afghan War
HBWS5	Iraq War
HD	Archaeology
HDD	Archaeology by period / region
HDDA	Prehistoric archaeology
HDL	Landscape archaeology
HDP	Environmental archaeology
HDR	Underwater archaeology
HDT	Industrial archaeology
HP	Philosophy
HPC	History of Western philosophy
HPCD	Western philosophy: c 1600 to c 1900
HPCD1	Western philosophy: Enlightenment
HPCF	Western philosophy, from c 1900 -
HR	Religion & beliefs
HRA	Religion: general
HRAB	Philosophy of religion
HRAF	Interfaith relations
HRAM	Religious issues & debates
HRAM1	Religious ethics

Code	Heading
HRAM2	Religion & politics
HRAM3	Religion & science
HRAM6	Religious fundamentalism
HRAM7	Blasphemy, heresy, apostasy
HRAM9	Religious intolerance, persecution & conflict
HRAX	History of religion
HRC	Christianity
HRCC	Christian Churches & denominations
HRCC1	The Early Church
HRCC2	Church history
HRCC7	Roman Catholicism, Roman Catholic Church
HRCC8	Orthodox & Oriental Churches
HRCC9	Protestantism & Protestant Churches
HRCC91	Anglican & Episcopalian Churches, Church of England
HRCC92	Baptist Churches
HRCC93	Calvinist, Reformed & Presbyterian Churches
HRCC95	Methodist Churches
HRCC96	Pentecostal Churches
HRCC97	Quakers (Religious Society of Friends)
HRCC99	Other Nonconformist & Evangelical Churches

Code	Heading
HRCJ	Ecumenism
HRCX	Christian institutions & organizations
HRCX1	Christian leaders & leadership
HRCX4	Christian ministry & pastoral activity
HRCX6	Christian social thought & activity
HRCX7	Christian mission & evangelism
HRCX8	Christian communities & monasticism
HRCZ	Christian & quasi-Christian cults & sects
HRE	Buddhism
HRG	Hinduism
HRH	Islam
HRJ	Judaism
HRK	Other non-Christian religions
HRKB	Baha'i
HRKJ	Jainism
HRKN	Oriental religions
HRKN1	Confucianism
HRKN3	Shintoism
HRKN5	Taoism
HRKS	Sikhism

Code	Heading
HRKT	Tribal religions
HRKZ	Zoroastrianism
HRLP	Religious institutions & organizations
HRLP1	Religious & spiritual leaders
HRLP5	Religious social & pastoral thought & activity
HRLP7	Religious communities & monasticism
HRQ	Alternative belief systems
HRQA	Humanist & secular alternatives to religion
J	Society & social sciences
JF	Society & culture: general
JFHF	Folklore, myths & legends
JFS	Social groups
JFSL	Ethnic studies
JFSL1	Ethnic minorities & multicultural studies
JFSR	Religious groups: social & cultural aspects
JFSR2	Islamic studies
JP	Politics & government
KJZ	History of specific companies / corporate history
TTVR	Traditional trades & skills
WTHM	Museum, historic sites, gallery & art guides

Code	Heading
WTLC	Classic travel writing
WTLP	Expeditions
WTM	Places & peoples: general & pictorial works
YNA	Art: general interest (Children's / Teenage)
YNC	Music: general interest (Children's / Teenage)
YNGL	Libraries, museums, schools (Children's / Teenage)

Relevant BIC qualifiers:

Code	Heading
1D	Europe
1DB	British Isles
1DBK	United Kingdom, Great Britain
1DBKE	England
1DBKEA	East Anglia
1DBKEAC	Cambridgeshire
1DBKEAL	Lincolnshire
1DBKEAN	Norfolk
1DBKEAS	Suffolk
1DBKEAX	Essex

Code	Heading
1DBKEM	Midlands
1DBKEMD	Derbyshire & Peak District
1DBKEMF	Herefordshire
1DBKEML	Leicestershire
1DBKEMM	Northamptonshire
1DBKEMN	Nottinghamshire
1DBKEMP	Shropshire
1DBKEMR	Rutland
1DBKEMS	Staffordshire
1DBKEMT	Worcestershire
1DBKEMW	Warwickshire, West Midlands
1DBKEN	North West England
1DBKENC	Cheshire
1DBKENL	Lancashire, Greater Manchester, Merseyside
1DBKENM	Cumbria & Lake District
1DBKES	South & South East England
1DBKESB	Berkshire
1DBKESD	Bedfordshire
1DBKESF	Oxfordshire
1DBKESH	Hampshire

Code	Heading
1DBKESK	Kent
1DBKESL	London, Greater London
1DBKESR	Surrey
1DBKEST	Hertfordshire
1DBKESU	Buckinghamshire
1DBKESW	Isle of Wight
1DBKESX	Sussex
1DBKEW	South West England
1DBKEWC	Cornwall
1DBKEWD	Devon
1DBKEWG	Gloucestershire
1DBKEWS	Somerset, Bristol
1DBKEWT	Dorset
1DBKEWW	Wiltshire
1DBKEY	North & North East England
1DBKEYD	Durham
1DBKEYK	Yorkshire
1DBKEYN	Northumberland, Tyne & Wear
1DBKN	Northern Ireland
1DBKS	Scotland

Code	Heading
1DBKSB	Lowland Scotland & Borders
1DBKSC	Central Scotland
1DBKSH	Northern Scotland, Highlands & Islands
1DBKSHF	Orkney Islands
1DBKSHJ	Shetland Islands
1DBKSHL	Western Isles, Outer Hebrides
1DBKSHQ	Isle of Skye
1DBKW	Wales
1DBKWC	Mid Wales
1DBKWN	North Wales
1DBKWS	South Wales
1DBKWV	Southwest Wales
1DBKX	Channel Islands
1DBKZ	Isle of Man
1DBR	Ireland
1DD	Western Continental Europe
1DDB	Belgium
1DDF	France
1DDFC	Corsica
1DDFM	Monaco

Code	Heading
1DDL	Luxembourg
1DDN	Netherlands
1DF	Central Europe
1DFA	Austria
1DFG	Germany
1DFGE	East Germany, DDR
1DFGW	West Germany
1DFH	Switzerland
1DFHA	Alps
1DFL	Liechtenstein
1DN	Northern Europe, Scandinavia
1DNC	Iceland
1DND	Denmark
1DNDF	Faroe Islands
1DNF	Finland
1DNN	Norway
1DNS	Sweden
1DS	Southern Europe
1DSE	Spain
1DSEB	Balearic islands

Code	Heading
1DSEP	Pyrenees
1DSG	Gibraltar
1DSM	Malta
1DSN	Andorra
1DSP	Portugal
1DST	Italy
1DSTA	Sardinia
1DSTC	Sicily
1DSU	San Marino
1DSV	Vatican
1DV	Eastern Europe
1DVC	Cyprus
1DVG	Greece
1DVGS	Greek islands
1DVGSC	Crete
1DVH	Hungary
1DVK	Former Czechoslovakia
1DVKC	Czech Republic
1DVKS	Slovakia
1DVP	Poland

Code	Heading
1DVT	Turkey
1DVU	Former Soviet Union, USSR (Europe)
1DVUA	Russia
1DVUAC	Chechnya
1DVUB	Belarus (Belorussia)
1DVUC	Latvia
1DVUE	Estonia
1DVUF	Lithuania
1DVUG	Georgia
1DVUK	Ukraine
1DVUM	Moldova (Moldavia)
1DVUR	Armenia
1DVUZ	Azerbaijan
1DVW	Southeast Europe
1DVWA	Albania
1DVWB	Bulgaria
1DVWR	Romania
1DVWY	Yugoslavia & former Yugoslavia
1DVWYB	Bosnia-Herzegovina
1DVWYC	Croatia

Code	Heading
1DVWYK	Kosovo
1DVWYM	Macedonia
1DVWYN	Montenegro
1DVWYS	Serbia
1DVWYV	Slovenia
1FBH	Israel
1QDAR	Ancient Rome
1QDH	Holy Roman Empire
1QDU	Austro-Hungarian Empire
1QFE	EU (European Union)
1QFW	Warsaw Pact, Eastern bloc
1QRM	Mediterranean countries
1QSB	Baltic Sea
1QSE	Irish Sea
1QSF	North Sea
1QSG	English Channel
1QSM	Mediterranean Sea

11.3 CCE

Note that in CCE there are many more differences from the current BIC headings than are listed here; they are not listed here since they describe topics that are not considered (directly) relevant to the current project.

Relevant CCE additional heading and geographic qualifiers:

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
HBJD2	Storia d'Italia	CCE only	x	
1DSTN	Italia Settentrionale	CCE only		x
1DSTNA	Valle d'Aosta	CCE only		x
1DSTNE	Piemonte	CCE only		x
1DSTNEA	Torino	CCE only		x
1DSTNEB	Alessandria	CCE only		x
1DSTNEC	Asti	CCE only		x
1DSTNED	Biella	CCE only		x
1DSTNEE	Cuneo	CCE only		x
1DSTNEF	Novara	CCE only		x
1DSTNEG	Provincia del Verbano Cusio Ossola	CCE only		x
1DSTNEH	Vercelli	CCE only		x
1DSTNG	Liguria	CCE only		x
1DSTNGA	Genova	CCE only		x
1DSTNGB	Imperia	CCE only		x
1DSTNGC	La Spezia	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTNGD	Savona	CCE only		x
1DSTNM	Lombardia	CCE only		x
1DSTNMA	Milano	CCE only		x
1DSTNMB	Bergamo	CCE only		x
1DSTNMC	Brescia	CCE only		x
1DSTNMD	Como	CCE only		x
1DSTNME	Cremona	CCE only		x
1DSTNMF	Lecco	CCE only		x
1DSTNMG	Lodi	CCE only		x
1DSTNMH	Mantova	CCE only		x
1DSTNMI	Monza e Brianza	CCE only		x
1DSTNMJ	Pavia	CCE only		x
1DSTNMK	Sondrio	CCE only		x
1DSTNML	Varese	CCE only		x
1DSTNN	Veneto	CCE only		x
1DSTNNA	Venezia	CCE only		x
1DSTNNB	Belluno	CCE only		x
1DSTNNC	Padova	CCE only		x
1DSTNND	Rovigo	CCE only		x
1DSTNNE	Treviso	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTNNF	Verona	CCE only		x
1DSTNNG	Vicenza	CCE only		x
1DSTNT	Trentino Alto Adige	CCE only		x
1DSTNTA	Bolzano	CCE only		x
1DSTNTB	Trento	CCE only		x
1DSTNU	Friuli Venezia Giulia	CCE only		x
1DSTNUA	Trieste	CCE only		x
1DSTNUB	Gorizia	CCE only		x
1DSTNUC	Pordenone	CCE only		x
1DSTNUD	Udine	CCE only		x
1DSTT	Italia Centrale	CCE only		x
1DSTTE	Emilia Romagna	CCE only		x
1DSTTEA	Bologna	CCE only		x
1DSTTEB	Ferrara	CCE only		x
1DSTTEC	Forlì-Cesena	CCE only		x
1DSTTED	Modena	CCE only		x
1DSTTEE	Parma	CCE only		x
1DSTTEF	Piacenza	CCE only		x
1DSTTEG	Ravenna	CCE only		x
1DSTTEH	Reggio nell'Emilia	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTTEI	Rimini	CCE only		x
1DSTTM	Marche	CCE only		x
1DSTTMA	Ancona	CCE only		x
1DSTTMB	Ascoli Piceno	CCE only		x
1DSTTMC	Fermo	CCE only		x
1DSTTMD	Macerata	CCE only		x
1DSTTME	Pesaro e Urbino	CCE only		x
1DSTTN	Toscana	CCE only		x
1DSTTNA	Firenze	CCE only		x
1DSTTNB	Arezzo	CCE only		x
1DSTTNC	Grosseto	CCE only		x
1DSTTND	Livorno	CCE only		x
1DSTTNE	Lucca	CCE only		x
1DSTTNF	Massa-Carrara	CCE only		x
1DSTTNG	Pisa	CCE only		x
1DSTTNH	Pistoia	CCE only		x
1DSTTNI	Prato	CCE only		x
1DSTTNJ	Siena	CCE only		x
1DSTTR	Umbria	CCE only		x
1DSTTRA	Perugia	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTTRB	Terni	CCE only		x
1DSTTZ	Lazio	CCE only		x
1DSTTZA	Roma	CCE only		x
1DSTTZB	Frosinone	CCE only		x
1DSTTZC	Latina	CCE only		x
1DSTTZD	Rieti	CCE only		x
1DSTTZE	Viterbo	CCE only		x
1DSTU	Italia Meridionale e Isole	CCE only		x
1DSTUA	Abruzzo	CCE only		x
1DSTUAA	L'Aquila	CCE only		x
1DSTUAB	Chieti	CCE only		x
1DSTUAC	Pescara	CCE only		x
1DSTUAD	Teramo	CCE only		x
1DSTUE	Molise	CCE only		x
1DSTUEA	Campobasso	CCE only		x
1DSTUEB	Isernia	CCE only		x
1DSTUM	Campania	CCE only		x
1DSTUMA	Napoli	CCE only		x
1DSTUMB	Avellino	CCE only		x
1DSTUMC	Benevento	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTUMD	Caserta	CCE only		x
1DSTUME	Salerno	CCE only		x
1DSTUP	Puglia	CCE only		x
1DSTUPA	Bari	CCE only		x
1DSTUPB	Barletta Andria Trani	CCE only		x
1DSTUPC	Brindisi	CCE only		x
1DSTUPD	Foggia	CCE only		x
1DSTUPE	Lecce	CCE only		x
1DSTUPF	Taranto	CCE only		x
1DSTUR	Calabria	CCE only		x
1DSTURA	Reggio Calabria	CCE only		x
1DSTURB	Catanzaro	CCE only		x
1DSTURC	Cosenza	CCE only		x
1DSTURD	Crotone	CCE only		x
1DSTURE	Vibo Valentia	CCE only		x
1DSTUS	Basilicata	CCE only		x
1DSTUSA	Potenza	CCE only		x
1DSTUSB	Matera	CCE only		x
1DSTUT	Sardegna	CCE only		x
1DSTUTA	Cagliari	CCE only		x

Code	Heading	CCE?	Directly relevant	Relevant as qualifier
1DSTUTB	Carbonia-Iglesias	CCE only		x
1DSTUTC	Provincia del Medio Campidano	CCE only		x
1DSTUTD	Nuoro	CCE only		x
1DSTUTE	Provincia dell'Ogliastra	CCE only		x
1DSTUTF	Olbia Tempio	CCE only		x
1DSTUTG	Oristano	CCE only		x
1DSTUTH	Sassari	CCE only		x
1DSTUV	Sicilia	CCE only		x
1DSTUVA	Palermo	CCE only		x
1DSTUVB	Agrigento	CCE only		x
1DSTUVC	Caltanissetta	CCE only		x
1DSTUVD	Catania	CCE only		x
1DSTUVE	Enna	CCE only		x
1DSTUVF	Messina	CCE only		x
1DSTUVG	Ragusa	CCE only		x
1DSTUVH	Siracusa	CCE only		x
1DSTUVI	Trapani	CCE only		x

11.4 WARENGRUPPEN 2.0

WG code	WG heading
1111	HC/Belletristik/Hauptwerk vor 1945
1114	HC/Belletristik/Märchen, Sagen, Legenden
1117	HC/Belletristik/Briefe, Tagebücher
1118	HC/Belletristik/Essays, Feuillton, Literaturkritik, Interviews
1150	HC/Belletristik/Lyrik, Dramatik
1151	HC/Belletristik/Lyrik
1152	HC/Belletristik/Dramatik
1230	HC/Kinder- und Jugendbücher/Vorlesebücher, Märchen, Sagen, Reime, Lieder
1232	HC/Kinder- und Jugendbücher/Märchen, Sagen
1233	HC/Kinder- und Jugendbücher/Gedichte, Reime
1234	HC/Kinder- und Jugendbücher/Lieder, Songs
1270	HC/Kinder- und Jugendbücher/Biografien
1351	HC/Reisen/Bildbände/Deutschland
1352	HC/Reisen/Bildbände/Europa
1361	HC/Reiseberichte, Reiseerzählungen/Deutschland
1362	HC/Reiseberichte, Reiseerzählungen/Europa
1522	HC/Philosophie/Antike
1523	HC/Philosophie/Mittelalter

WG code	WG heading
1524	HC/Philosophie/Renaissance, Aufklärung
1525	HC/Philosophie/Deutscher Idealismus, 19. Jahrhundert
1553	HC/Geschichte/Altertum
1554	HC/Geschichte/Mittelalter
1555	HC/Geschichte/Neuzeit bis 1918
1558	HC/Geschichte/Regionalgeschichte, Ländergeschichte
1559	HC/Geschichte/Kulturgeschichte
1582	HC/Kunst/Kunstgeschichte
1583	HC/Kunst/Bildende Kunst
1584	HC/Kunst/Architektur
1585	HC/Kunst/Innenarchitektur, Design
1586	HC/Kunst/Theater, Ballett
1587	HC/Kunst/Fotografie, Film, Video, TV
1588	HC/Kunst/Antiquitäten
1593	HC/Musik/Musikgeschichte
1941	HC/Sachbücher/Geschichte/Biographien, Autobiographien
1943	HC/Sachbücher/Geschichte/Regionalgeschichte, Ländergeschichte
1945	HC/Sachbücher/Geschichte/Mittelalter
1946	HC/Sachbücher/Geschichte/Neuzeit bis 1918
1951	HC/Sachbücher/Kunst, Literatur/Biographien, Autobiographien

WG code	WG heading
1953	HC/Sachbücher/Kunst, Literatur/Bildende Kunst
1954	HC/Sachbücher/Kunst, Literatur/Fotokunst
1955	HC/Sachbücher/Kunst, Literatur/Architektur
1956	HC/Sachbücher/Kunst, Literatur/Literatur/Allgemeines, Nachschlagewerke
1961	HC/Sachbücher/Musik, Film, Theater/Biographien, Autobiographien
1963	HC/Sachbücher/Musik, Film, Theater/Klassik, Oper, Operette, Musical
1964	HC/Sachbücher/Musik, Film, Theater/Jazz, Blues
1965	HC/Sachbücher/Musik, Film, Theater/Pop, Rock
1966	HC/Sachbücher/Musik, Film, Theater/Film/Allgemeines, Nachschlagewerke
1967	HC/Sachbücher/Musik, Film, Theater/TV/Allgemeines, Nachschlagewerke
1968	HC/Sachbücher/Musik, Film, Theater/Theater, Ballett/Allgemeines, Nachschlagewerke

11.5 DEWEY DECIMAL

Dewey Decimal Classification subject headings which are relevant for Linked Heritage, taken, not from the whole DDC scheme, but rather from the subset of headings used in the actual book data provided by the e-book publisher / retailer in case study 4.5.5. Also included for interest are the frequencies of occurrence of each subject code within the relevant subset. The sorting by frequency of occurrence (most frequent to least) also demonstrates that the most general subject headings are used most often as would be expected from the considerations in section 4.4.12.

Dewey subject heading	Dewey code	Frequency
Italian fiction	853	438
General history of Europe Italian Peninsula & adjacent islands	945	409

Italian poetry	851	331
Italian drama	852	200
General history of Europe	940	140
Spanish & Portuguese literatures	860	92
History of ancient world Italy & adjacent territories	937	78
General history of Europe Iberian Peninsula & adjacent islands	946	69
Roman Catholic Church	282	57
Modern Western philosophy Germany & Austria	193	55
Europe	914	51
Modern Western philosophy Italy	195	49
Spanish fiction	863	41
Modern Western philosophy France	194	34
Christian church history	270	26
French fiction	843	23
English fiction	823	21
Galleries, museums, private collections	708	21
Spanish drama	862	21
Spanish poetry	861	21
General history of Europe Other parts of Europe	949	19
English & Old English literatures	820	17
History of ancient world Greece	938	16

French drama	842	15
Italic literatures Latin	870	15
German fiction	833	12
Latin poetry	871	12
News media, journalism, publishing In Italy & adjacent territories	075	11
English drama	822	10
French poetry	841	10
General history of Europe Eastern Europe Soviet Union	947	10
Modern Western philosophy Spain & Portugal	196	10
Modern Western philosophy British Isles	192	9
General history of Europe Central Europe Germany	943	8
German poetry	831	8
History of ancient world Europe north & west of Italy	936	8
Medieval Western philosophy	189	8
Architecture to ca. 300	722	7
Classical (Greek & Roman) religion	292	7
Gospels & Acts	226	7
English poetry	821	6
General history of Europe England & Wales	942	6
Protestants of Continental origin	284	6
Classical Greek epic poetry & fiction	883	5

General history of Europe France & Monaco	944	5
Classical Greek drama	882	4
Greek, Etruscan, Roman sculpture	733	4
Italian speeches	855	4
German drama	832	3
Italian essays	854	3
Ancient world	913	2
Chamber music	785	2
Classical Greek poetry	881	2
General collections In Italian, Romanian, Rhaeto-Romanic	085	2
General history of Europe British Isles	941	2
Latin dramatic poetry & drama	872	2
Latin epic poetry & fiction	873	2
Modern Western philosophy Scandinavia	198	2
New Testament	225	2
Pre-Socratic Greek philosophies	182	2
Architecture from ca. 300 to 1399	723	1
French essays	844	1
General history of Europe Northern Europe Scandinavia	948	1
Modern Western philosophy Soviet Union	197	1
Revelation (Apocalypse)	228	1

Spanish essays	864	1

12 APPENDIX – TEXT OF NEW COOPERATION AGREEMENT FOR COMMERCIAL SECTOR ("PROTOTYPE") DATA CONTRIBUTORS

Linked Heritage

Commercial Data Contribution Authorisation Agreement

A cooperation agreement is hereby concluded between:

Ministero per i beni e le attività culturali - Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, Viale Castro Pretorio 105, 00185 Roma in its role of Coordinator of the Linked Heritage project ("The Coordinator")

of the one part,

and

EDItEUR, Ltd., United House, North Road, London N7 9DP ("EDItEUR")

of the other part,

collectively "the Parties", represented by their authorised representatives.

This agreement is to be interpreted under the law of Italy.

In the framework of the project "Linked Heritage" – Coordination of Standards and Technologies for the enrichment of Europeana" funded by ICT Policy Support Programme, Grant Agreement no. 270905.

1. **Terms**

    a. **Commercial Data Contribution Authorisation Agreement**

This agreement, made between The Coordinator and EDItEUR, within the framework of the existing agreements and contracts governing Linked Heritage.

    b. **The Coordinator – as defined above.**

    c. **EDItEUR – as defined above.**

    d. **The Parties – as defined above.**

    e. **Linked Heritage**

The 30-month EC-funded project, beginning 1st April 2011. CIP Best Practice Network, grant agreement 270905, in the framework of ICT-PSP, 7th Framework.

    f. **Data Contribution**

A set of metadata records in XML format made available to EDItEUR under the terms set out in the Commercial Data Contribution Authorisation Agreement and Commercial Data Contributor agreement appended thereto.

    g. **MINT**

The data aggregation software application used by EDItEUR within Linked Heritage to store Data Contributions, convert their XML schema to LIDO for internal technical processes, and pass the Publication Subset to Europeana after transforming it to ESE/EDM.

h. LIDO

The XML data format used internally to Linked Heritage for storage and transformation of data prior to publication to Europeana.

i. Europeana

The Europeana Foundation (Koninklijke Bibliotheek, Prins Willem-Alexanderhof 5, 2509 The Hague, The Netherlands), its www.europeana.eu portal database, Linked Open Data dumps and APIs, taken collectively.

j. ESE/EDM

Collectively, ESE and/or EDM; the data formats accepted by Europeana. The formats in which the Publication Subset will be delivered by EDItEUR to Europeana after transformation from the LIDO format using MINT.

k. Commercial Data Contributor Agreement

The terms and conditions (in Annex 1) under which Contributors will pass Data Contributions to EDItEUR for use within Linked Heritage.

l. Contributor

Commercial organisation who passes Data Contributions to EDItEUR for use within Linked Heritage under the terms specified in the Commercial Data Contributor Agreement.

m. Publication Subset

The subset of data elements of each Record within the Data Contribution that, by under a Commercial Data Contributor Agreement, EDItEUR passes to Europeana.

n. DEA

The terms and conditions ("Data Exchange Agreement") under which EDItEUR will pass the Publication Subset of Data Contributions to Europeana. This gives Europeana the right to re-publish the Publication Subset under the Creative Commons CC0 framework.The addition of links to Thumbnail Previews by Europeana upon receipt of the Publication Subset is Europeana's standard procedure but can be omitted on instruction by EDItEUR at the Contributor's request.

o. Prototype Data

A Data Contribution for which the Contributor has specified a Publication Subset to be passed to Europeana by EDItEUR. Whether a Data Contribution is Prototype Data or only Test Data must be specified in the Technical Annex for each separate Data Contribution.

p. Test Data

A Data Contribution for which the Contributor has specified that no Publication Subset shall be passed to Europeana. Test Data will only be used internally to Linked Heritage.

2. Subject matter and scope

1. EDItEUR is a partner with The Coordinator in the Linked Heritage project with specific rights and duties laid out in the Linked Heritage Grant Agreement and Description of

Work. This agreement does not alter or add to these in any way. It is limited only to specifying how two of these activities will be carried out, namely:

a) receiving Data Contributions from commercial Contributors and uploading them to MINT, and transforming the Data Contributions into LIDO format;

b) signature of the Europeana DEA in respect of the Publication Subset of the Data Contributions, and passing the Publication Subset to Europeana through MINT, in ESE/EDM format.

2. The Coordinator authorises EDItEUR, Ltd. to enter into Commercial Data Contributor Agreements defined in Annex 1 with other organisations, in the framework of the linked Heritage project and for the purposes described above.
3. The specific uses to which the Data will be put and the limits on this use are described exhaustively in the Commercial Data Contributor agreement in Annex 1.
4. This agreement applies only in the context of MINT and Europeana, and its term is limited to that of the Linked Heritage project.
5. No monetary compensation for either of The Parties is implied by this agreement.
6. The Parties hereby guarantee that they are able to grant all the rights listed below.
7. Unless otherwise stated herein, the Parties have no warrant to assign any these rights to any other party.

3. Rights and duties

1) *EDItEUR* shall have the right to enter into Commercial Data Contributor agreements (specified in Annex 1) with commercial organisations, for the following, and no other, purposes:

a. Receiving and ingesting Data Contributions from Contributors to MINT;

b. Using the Data Contributions (both Test Data and Prototype Data) in MINT only in order to:

i. test the LIDO schema mappings developed under Linked Heritage Work Package 4;

ii. pass the Publication Subset of any Data Contribution marked as Prototype Data to Europeana;

2) *The Coordinator* shall have the duty to:

a. Ensure the security of the Data uploaded by EDItEUR into the Linked Heritage MINT server, including the following guarantees;

i. No part of the Data Contribution will be released outside of MINT in any format, except the Publication Subset passed on to Europeana;

ii. The Contributor shall be guaranteed access to the Data Contribution for the duration of the agreement, including the right to correct or remove the Data Contribution from MINT (such correction to be effective within 10 business days);

iii. All Data Contributions, in whatever format, and in all copies and versions, shall be removed entirely from MINT within 30 days of expiry or termination of this agreement.

4. Entry into force of the Commercial Data Contribution Authorisation Agreement

The Cooperation Agreement shall enter into force from the date of its signature.

5. **Completion, expiry or termination of the Commercial Data Contribution Authorisation Agreement**

- The present Commercial Data Contribution Authorisation Agreement shall expire on the date of completion or termination of the Linked Heritage Grant Agreement with the European Commission, or on 31/12/2013 (whichever is the earliest).
- Commercial Data Contributor Agreements made under this original Commercial Data Contribution Authorisation Agreement shall also expire on the date of completion or termination of the Linked Heritage Grant Agreement with the European Commission, or on 31/12/2013 (whichever is the earliest), and may not be presumed to continue after that, even should this Commercial Data Contribution Authorisation Agreement be renewed.
- In the case that the Linked Heritage project will continue its activities beyond the completion of the Grant Agreement with the European Commission:
- this Commercial Data Contribution Authorisation Agreement can be renewed by mutual agreement of the Parties;
- any Commercial Data Contributor Agreement made under the original Commercial Data Contribution Authorisation Agreement must be renewed directly with each individual Data Contributor.
- Either party may terminate this agreement at any time and for any reason, subject to one month's written notice.
- Notwithtanding the above, The Coordinator's duties in 2.2 shall survive expiry or termination of this agreement.

For the Parties

The Coordinator

Ministero per i Beni e le Attività Culturali

Istituto centrale per il catalogo unico delle biblioteche italiane

Viale Castro Pretorio 105

00185 Roma, Italy

NAME OF THE AUTHORISED REPRESENTATIVE: Rossella Caffo

TITLE: Director

SIGNATURE:

DATE:

EDItEUR

EDItEUR, Ltd., United House, North Road, London N7 9DP

NAME OF THE AUTHORISED REPRESENTATIVE: XXX

TITLE: XXX

SIGNATURE:

DATE:

Annex 1

Linked Heritage

Commercial Data Contributor Agreement

A cooperation agreement is hereby concluded between:

EDItEUR, Ltd. (“EDItEUR”)

United House, North Road, London N7 9DP,

of the one part,

and

Name, Address (“The Contributor”)

of the other part,

collectively “the Parties”, represented by their authorised representatives.

This agreement is to be interpreted under the law of England and Wales.

In the framework of the Linked Heritage project and within the “Commercial Data Contribution Authorisation Agreement” signed by ICCU and EDItEUR.

The Parties have agreed to cooperate as follows in order to contribute data via EDItEUR to Linked Heritage.

1. **Terms**

a. Commercial Data Contributor Agreement

This form of agreement.

b. Contributor

The commercial organisation entering into this agreement with EDItEUR.

c. EDItEUR – as defined above.

d. The Parties – as defined above.

e. Linked Heritage

The 30-month EC-funded project, beginning 1st April 2011. CIP Best Practice Network, grant agreement 270905, in the framework of ICT-PSP, 7th Framework.

f. Data Contribution

A set of metadata Records in XML format made available to EDItEUR under the terms set out in this Commercial Data Contributor agreement.

g. MINT

The data aggregation software application used by EDItEUR within Linked Heritage to store Data Contributions, convert their XML schema to LIDO for internal technical processes, and pass the Publication Subset to Europeana.

h. LIDO

The XML data format used internally to Linked Heritage for storage and transformation of data prior to publication to Europeana.

i. Europeana

The Europeana Foundation (Koninklijke Bibliotheek, Prins Willem-Alexanderhof 5, 2509 The Hague, The Netherlands), its www.europeana.eu portal database, Linked Open Data dumps and APIs, taken collectively.

j. Commercial Data Contributor Agreement

The terms and conditions under which commercial data contributors will pass Data Contributions to EDItEUR for use within Linked Heritage.

k. Publication Subset

The subset of data, defined in the Technical Annex by a listing of the data elements of each Record within the Data Contribution, that, by arrangement with EDItEUR under a Commercial Data Contributor Agreement, EDItEUR passes to Europeana under the DEA.

l. DEA

The terms and conditions under which EDItEUR will pass the Publication Subset of Data Contributions to Europeana.

m. Digital Image File

The highest resolution digital image file representing a Commercial Product; normally (but not always) found at the Retail Web Page for that Commercial Product. A Digital Image File's URL (one per Commercial Product) is an optional element of the Publication Subset; if it is contributed, Europeana will use it to generate a Thumbnail Preview.

n. Thumbnail Previews

Low resolution previews generated by Europeana from any available Digital Image File for use in its www.europeana.eu portal, and hosted at Europeana's own servers. Their URL (*in addition to* that of the source Digital Image File) is added to the Publication Subset by Europeana and released under CC0 by Europeana (unless the omission of the Thumbnail Preview URL is requested in the Technical Annex).

o. Commercial Product

A commercially available, copyrighted creation available for retail purchase in a specific form from the Contributor and described by a Record within the Data Contribution.

p. Record

An XML document containing metadata, part of the Data Contribution, describing one, and only one, Commercial Product (although a Retail Web Page *may* describe various related Commercial Products, *e.g.* differing in form but derived from the same source work, or identical in form but in different digital and physical formats).

q. Retail Web Page

A Web page, whose URL is provided in the Record for a Commercial Product, where that Commercial Product can be bought via ecommerce by the general public.

r. Prototype Data

A Data Contribution for which the Contributor has specified a Publication Subset to be passed to Europeana by EDItEUR, upon final confirmation in writing to EDItEUR. Whether a Data Contribution is Prototype Data or only Test Data must be specified in the Technical Annex for each separate Data Contribution.

s. Test Data

A Data Contribution for which the Contributor has specified that no Publication Subset shall be passed to Europeana. Test Data will only be used internally to Linked Heritage.

2. Subject matter and scope

1. This agreement specifies how commercial sector organisations may:
   a) deliver Data Contributions to EDItEUR, for upload to MINT and translation to the LIDO metadata format ("LIDO");
   b) give EDItEUR the right to pass the Publication Subset, agreed with the Contributor as in the Technical Annex, to Europeana through MINT as ESE/EDM data.
2. This is a non-exclusive agreement.
3. It applies in the context of the MINT and Europeana applications, and its term is limited to that of the Linked Heritage project.
4. No monetary compensation for either of The Parties is implied by this agreement.
5. The Parties hereby warrant that they are empowered to grant all the rights listed below;
   a. For the avoidance of doubt, the right to distribute a URL shall not imply any rights over the content (in any form) located at that URL;
   b. The Contributor retains all existing rights in Data Contributions (and their constitutive Records) they contribute to EDItEUR (except in the Publication Subset once it is passed to Europeana)
6. Note that the Publication Subset of Data Contributions designated as Prototype Data may have been provided to Europeana under CC0 terms, and this agreement has no bearing on that subset once passed to Europeana.

3. Rights and duties

The *Contributor* will have the rights to:

1. Request the removal of any or all of their Data Contributions (in whatever format, as above) from MINT at any time;
2. Designate in writing which Data Contributions are Test Data or Prototype Data, and for Prototype Data,
   a. specify which data elements will make up the Publication Subset, using the form enclosed here as the Technical Annex;
   b. request omission, if desired, of the URLs of Thumbnail Previews from Europeana's CC0 data releases;
3. Resubmit the Technical Annex in order to change the status of specific Data Contributions, previously designated as Test Data, to make them Prototype Data by specifying a Publication Subset for them.

The *Contributor* will have the duties to:

1. Provide Data Contributions detailed in the annex to this agreement;

2. Warrant that all data contained in the Publication Subset of any and each Data Contribution designated as Prototype Data has been cleared for release by Europeana under the terms of the DEA, principally the Creative Commons CC0 terms.

EDItEUR will have the duties to:

1. Ensure the security of all Data Contributions including guaranteeing that no part of the Data Contribution will be released outside of MINT in any format, except the Publication Subset passed on to Europeana;
2. Provide copies of the original Data Contribution or its translated versions in LIDO or ESE/EDM, within 10 business days of request, for the duration of the agreement
3. Correct or remove the Data from MINT within one or two business days of request by the Contributor;
4. Completely delete from MINT all Data Contributions, in whatever format, and in all copies and versions, upon expiry or termination of this agreement.

EDItEUR will have the rights to:

1. Upload the Data Contributions into MINT for the purposes of creating, testing, and demonstrating technical schema mappings to LIDO and other research purposes;
    a. Grant *limited* rights to other members of the Linked Heritage consortium to view the data and edit schema mappings *solely* within MINT;
2. Provide the Publication Subset of Data Contributions specified in the Technical Annex, and designated therein as Prototype Data to Europeana;
3. Grant Europeana the right to generate, store and display a Preview Thumbnail created from each Digital Image File for which the Contributor includes a Digital Image File URL in a Publication Subset.
    b. For the avoidance of doubt, EDItEUR does not grant Europeana any further rights over the Digital Image File or Preview Thumbnail derived therefrom.

Article 3 - Entry into force of the Commercial Data Contributor Agreement

The Cooperation Agreement shall enter into force from the date of its signature.

Article 4 - Completion, expiry or termination of the Commercial Data Contributor Agreement

- The present Cooperation Agreement shall terminate on the date of completion or termination of the Linked Heritage Grant Agreement with the European Commission or 31/12/2013 (whichever is the earliest).
- In the case that the Linked Heritage project will continue its activities beyond the completion of the Grant Agreement with the European Commission, this Commercial Data Contributor Agreement can be renewed by mutual agreement of the Parties.
- Each party may terminate this agreement at any time and for any reason, subject to one month's written notice.

For the Parties

EDItEUR

EDItEUR, Ltd., United House, North Road, London N7 9DP

NAME OF THE AUTHORISED REPRESENTATIVE: XXX

TITLE: XXX

SIGNATURE:

DATE:

The Contributor

ORGANISATION: XXX

ADDRESS: XXX

NAME OF THE AUTHORISED REPRESENTATIVE: XXX

TITLE: XXX

SIGNATURE:

DATE:

Technical Annex

the Contributor shall provide Data Contributions in the form of the following XML Records in the standard industry formats indicated[137]:

Data Contribution #[138]	**Provider[139]**	**Filenames of data records**	**Subject matter of the data supplied**	**Language of metadata**	**Format[140]**	**IPR[141]**
			Books		ONIX for Books (version 2.1 or 3.0?)	
			Film and TV		EIDR	
			Music		DDex (version?)	
			Photos		IPTC/XMP	

[137] Linked Heritage will only accept XML metadata in the standard formats listed here for each commercial product media type.

[138] Each separate Data Contribution is numbered within this agreement to link it with a Publication Subset definition (see next pages).

[139] Name and contact details of any relevant third party who provides a subset of the Data Contribution.

[140] Source schema of XML file, or for Digital Image Files, format of digital content (file format, resolution etc) referred to by URLs provided (actual image files need *not* be provided).

[141] Is the partner the owner of the IPR in the metadata (or the image file at the URL supplied)? If not, which agreement the partner has with the IPR owner? Or, is the XML or image file publicly available?

Data Contribution #[138]	**Provider**[139]	**Filenames of data records**	**Subject matter of the data supplied**	**Language of metadata**	**Format**[140]	**IPR**[141]
[reference # of Data Contribution for which these are the Digital Image File URLs]			URLs (only) of Digital Image Files [one row per Data Contribution above]		[format of image files]	

Publication Subset definition

Insert a new Publication Subset definition form below for each Data Contribution listed above.

The first 8 items (A-H) *require* a response from the Contributor.

A	This is the Publication Subset for Data Contribution number…		1 [change to the correct number for subsequent forms]
B	I am providing… [indicate TEST data, and insert "OMIT" and "NO" in all rows above if you do not wish any Publication Subset to be passed to Europeana. For PROTOTYPE data the Publication Subset you define below will be released to Europeana]		[TEST] / [PROTOTYPE] DATA [delete as applicable] [to change the status of a Data Contribution previously designated TEST DATA, re-submit this form using the reference number for that Data Contribution, changing this box to PROTOTYPE DATA and specifying the mandatory fields below]
Please indicate below the metadata fields* that may be included in the Publication Subset.			
C	**At least one of (either/or)**	Commercial Product's title	YES [recommended]
D		Free-text description of Commercial Product	NO [default value]
URLs of Digital Image Files for generation of Thumbnail Previews: • The Publication Subset *does not and cannot* include Digital Image Files themselves. • The inclusion of a URL showing a retail page for each product (one link per product) is *strongly recommended.* • *Either* the Digital Image File URL *or* the Retail Web Page URL *must* be included.			
E	**At least one of (either/or)**	URL of Digital Image file for Thumbnail Preview generation	NO [default value]
F		URL of Retail Web Page	YES [recommended]

Redistribution of Thumbnail Previews' URLs by Europeana: • Thumbnail Previews are generated internally by Europeana from Digital Image Files, and hosted and displayed at the www.europeana.eu portal together with a rights reserved statement (www.europeana.eu/rights/rr-p.html). • The Thumbnail Previews themselves are *not* redistributed by Europeana. • Europeana *may* add *the URLs for its Thumbnail Previews* to the Publication Subset before redistributing this textual data under CC0 (http://creativecommons.org/publicdomain/zero/1.0/) but the addition of Thumbnail Preview URLs *is not* mandatory – specify below if this is to be omitted.		
G	State "OMIT" in this row to omit the URLs of the Thumbnail Previews hosted at Europeana from publication by Europeana under CC0	OMIT [default value]
H	*At least one more* field describing context of the Product (*e.g.* subject term, keyword, date, location…)	PUBLICATION DATE [suggested value]
I	[insert more rows for optional descriptive fields if desired]	

*Specific XML data fields that provide each of the above must be agreed by EDItEUR with the Contributor before a Publication Subset may be passed to Europeana.

13 APPENDIX – RECIPE FOR CREATING BASIC ONIX 3.0 RECORDS FROM SPREADSHEET DATA

1. Download BookNet Canada BNC Bronze template from:

   https://booknetcanada.atlassian.net/wiki/display/UserDocs/BNC+Bronze+Template [142]

2. Set code list values to equivalents from ONIX 2.1 code lists, if applicable

3. Insert placeholder data in mandatory (yellow header) columns if no real value available

4. Paste in copied values to BNC ONIX converter:

   http://booknet.gogpg.com/Converter.aspx

5. Set correct values for:
   a. AddresseeName = “LINKED HERITAGE”
   b. ExtentType – when created as placeholder for different Extent Types (e.g. e-book file size produced from audiobook running time in template); note that zeroes from the audiobook running time (in HHMMSS) may need to be deleted without adding whitespace into the element content

6. Delete (unless real data available):
   a. DefaultLanguageOfText
   b. DefaultPriceType
   c. Measure (and all subelements)
   d. SalesRightsType
   e. PriceType
   f. DiscountCodeType

7. Add extra fields manually according to needs of data supplier

[142] Note that this template is suitable only for exploratory work. It is not intended (even within Canada) for large-scale management of bibliographic data, and there may be restrictions on the use of the BNC ONIX Converter in step 4