



ICT-PSP Project no. 270905

LINKED HERITAGE

Coordination of standard and technologies
for the enrichment of Europeana

Starting date: 1st April 2011

Ending date: 31st October 2013

Deliverable Number:	D 4.1
Title of the Deliverable:	Best Practice Report – Public Private Partnership
Dissemination Level:	Public

Contractual Date of Delivery to EC:	March 2011
Actual Date of Delivery to EC:	December 2011

Project Co-ordinator

Company name : Istituto Centrale per il Catalogo Unico (ICCU)
Name of representative : Rosa Caffo
Address : Viale Castro Pretorio 105, I-00185 Roma
Phone number: +39.06.49210427
Fax number: +39.06. 06 4959302
E-mail: rcaffo@beniculturali.it
Project WEB site address: <http://www.linkedheritage.eu>

Context

WP 4	Public-Private Partnership
WP Leader	Michael Hopwood / EDItEUR
Task 4.1	Private Sector Business Exploration
Task Leader	Michael Hopwood / EDItEUR
Dependencies	

Author(s)	Michael Hopwood
Contributor(s)	Graham Bell
Reviewers	Collections Trust
Approved by:	

History

Version	Date	Author	Comments
0.1	06 September 2011	Michael Hopwood	First draft (1.32) internally reviewed by Graham Bell
0.2	19 October 2011	Michael Hopwood	Informal review draft (1.49) circulated to Work Group 4 for comment and revision
0.9	18 November 2011	Michael Hopwood	Formal review draft (2.3) submitted to Antonella Fresa and Gordon McKenna, and circulated to WG4
1.0	19 December 2011	Michael Hopwood	Final version

CONTENTS

1	EXECUTIVE SUMMARY	6
2	INTRODUCTION.....	8
2.1	LINKED HERITAGE, WORK PACKAGE 4	9
2.2	BACKGROUND.....	10
2.3	OVERVIEW	11
2.3.1	<i>The aim and originality of the research.....</i>	<i>11</i>
2.3.2	<i>Scope of the research and contents of the report</i>	<i>12</i>
2.3.3	<i>Limits of the project scope (what is not in this report).....</i>	<i>13</i>
2.4	LINKS WITH OTHER PROJECTS AND WORK PACKAGES	13
3	EUROPEANA DATA MODEL AND COMMERCIAL CONTENT.....	15
3.1	THE KIND OF CONTENT REQUIRED BY EUROPEANA	15
3.2	THE RICH METADATA FOUND IN THE COMMERCIAL SECTOR.....	16
3.3	SEMANTIC ENRICHMENT BENEFITS FROM COMMERCIAL METADATA.....	17
3.4	LIMITATIONS OF EUROPEANA TOP-LEVEL AGGREGATION	18
4	RESEARCH METHODS AND APPROACH	21
4.1	INDUSTRY CONTACTS AND BEST PRACTICE NETWORK	21
4.2	DESK RESEARCH	21
4.3	CONCEPTUAL FRAMEWORKS	22
5	COMMERCIAL SECTOR METADATA LANDSCAPE.....	24
5.1	EUROPEAN CULTURAL INDUSTRIES	24
5.1.1	<i>Size and Scope of the Cultural and Creative Economy in the EU.....</i>	<i>24</i>
5.1.2	<i>Correct scope of “cultural” or “content” industries.....</i>	<i>25</i>
5.1.3	<i>Cultural Industry sectors in detail.....</i>	<i>25</i>
5.1.4	<i>Modes of production and classification of products</i>	<i>26</i>
5.2	SUPPLY CHAINS	28
5.2.1	<i>Supply chains, identifiers and metadata schemas</i>	<i>28</i>
5.2.2	<i>Types of cultural industry businesses</i>	<i>29</i>
5.3	COMMERCIAL SECTOR DISTINCTIVES.....	30
5.3.1	<i>What is metadata?.....</i>	<i>30</i>
5.3.2	<i>Resource or copyright content focus?</i>	<i>31</i>
5.3.3	<i>How metadata is used in public cultural heritage sectors</i>	<i>32</i>
5.3.4	<i>How metadata is used in the commercial cultural industries</i>	<i>32</i>
5.3.5	<i>Metadata best practice</i>	<i>34</i>
5.3.6	<i>Standards: Origin, adoption and maintenance</i>	<i>36</i>
5.3.7	<i>Commercial sector identifiers.....</i>	<i>37</i>
5.4	CONCEPTUAL MODELS FOR DATA ARCHITECTURE AND INTEROPERABILITY	40
5.4.1	<i>CIDOC Conceptual Reference Model</i>	<i>41</i>
5.4.2	<i>FRBR(oo).....</i>	<i>42</i>
5.4.3	<i>Indecs Metadata Framework</i>	<i>44</i>
5.5	IDENTIFIERS AND METADATA FRAMEWORKS	45
5.5.1	<i>Overview of identifiers used in the cultural industries</i>	<i>47</i>
5.5.2	<i>Overview of metadata schemas used in the cultural industries.....</i>	<i>49</i>

6	INDUSTRY SECTOR FINDINGS	51
6.1	PRESENTATION OF RESULTS.....	51
6.2	CROSS-SECTORAL STANDARDS	52
6.2.1	<i>ISNI – International Standard Name Identifier.....</i>	<i>53</i>
6.2.2	<i>Other areas of cooperation and overlap</i>	<i>54</i>
6.3	BOOKS	55
6.3.1	<i>The book publishing sector.....</i>	<i>55</i>
6.3.2	<i>ISTC – International Standard Text Code.....</i>	<i>56</i>
6.3.3	<i>ISBN – International Standard Book Number.....</i>	<i>58</i>
6.3.4	<i>ISMN – International Standard Music Number</i>	<i>61</i>
6.3.5	<i>ISBN-A – Actionable ISBN</i>	<i>63</i>
6.3.6	<i>ONIX for Books</i>	<i>64</i>
6.4	MUSIC AND SOUND	68
6.4.1	<i>Music industry supply chains.....</i>	<i>68</i>
6.4.2	<i>ISWC – International Standard Musical Work Code.....</i>	<i>72</i>
6.4.3	<i>ISRC – International Standard Recording Code</i>	<i>73</i>
6.4.4	<i>GRIID – Global Release Identifier.....</i>	<i>75</i>
6.4.5	<i>DDEX Release Notification.....</i>	<i>79</i>
6.5	FILM AND TELEVISION.....	82
6.5.1	<i>The Audiovisual Sector</i>	<i>82</i>
6.5.2	<i>EIDR – Entertainment Identifier Registry.....</i>	<i>85</i>
6.5.3	<i>ISAN – International Standard Audiovisual Number</i>	<i>92</i>
6.5.4	<i>Technical standards and DRM: MPEG-7 and MPEG-21.....</i>	<i>95</i>
6.5.5	<i>Descriptive and marketing metadata.....</i>	<i>95</i>
6.6	PHOTOGRAPHY	96
6.6.1	<i>The photography sector</i>	<i>96</i>
6.6.2	<i>Outline workflows in photography industries</i>	<i>97</i>
6.6.3	<i>Rights and licensing in and of photos.....</i>	<i>100</i>
6.6.4	<i>Technical creation and reproduction of photo metadata</i>	<i>100</i>
6.6.5	<i>Identifiers for digital photographs?.....</i>	<i>102</i>
6.6.6	<i>IPTC and XMP</i>	<i>102</i>
7	LEGAL-COMMERCIAL IMPLICATIONS OF METADATA RE-USE.....	108
7.1	DRIVE FOR OPENNESS OF PUBLIC SECTOR METADATA	108
7.2	LEGAL STATUS OF COMMERCIAL METADATA	108
7.2.1	<i>Metadata normally make up databases</i>	<i>108</i>
7.2.2	<i>Databases are commercially significant.....</i>	<i>109</i>
7.2.3	<i>Confidentiality and implied licenses</i>	<i>109</i>
7.2.4	<i>Territorial sales rights and windowing</i>	<i>110</i>
8	EXISTING PARTNERSHIPS.....	111
8.1	PARTNERSHIP APPROACH ASSUMED IN LINKED HERITAGE.....	111
8.2	PROOFS OF CONCEPT	111
8.2.1	<i>Books: Gallica 2 (France).....</i>	<i>111</i>
8.2.2	<i>Books: Enclave (Spain).....</i>	<i>111</i>
8.2.3	<i>Film & TV: Ximon (Netherlands).....</i>	<i>112</i>
8.2.4	<i>Film & TV: FindAnyFilm.com.....</i>	<i>112</i>
9	CONCLUSIONS.....	114

9.1	COMMERCIAL METADATA IN GENERAL	114
9.2	SECTOR SUMMARY.....	115
9.2.1	<i>Cross-sector standards</i>	115
9.2.2	<i>Books</i>	115
9.2.3	<i>Music and sound</i>	116
9.2.4	<i>Film and TV</i>	116
9.2.5	<i>Photography</i>	117
9.3	PUBLIC-PRIVATE PARTNERSHIPS	117
9.4	IMPACT	117
9.4.1	<i>Contribution to Work Package 4</i>	118
9.4.2	<i>Relevance to other Working Groups</i>	118
9.4.3	<i>Broader relevance</i>	118
9.5	RECOMMENDATIONS AND WORK PLAN	118
9.5.1	<i>Recommendations for Linked Heritage</i>	118
9.5.2	<i>Work Plan</i>	119
10	ACKNOWLEDGEMENTS	121
11	REFERENCES.....	122
12	APPENDIX 1 – GLOSSARY OF TERMS	128
13	APPENDIX 2 – ACRONYMS USED.....	155
14	APPENDIX 3 – SAMPLE ONIX RECORDS	157
15	APPENDIX 4 – COMPARISON OF PHOTO METADATA FORMATS	169
16	APPENDIX 5 – SAMPLE EXIF, IPTC AND XMP RECORDS	174
17	APPENDIX 6 – LIST OF PICTURE LIBRARIES SURVEYED	191
18	APPENDIX 7 – SAMPLE DESCRIPTION OF IMPLIED LICENCE.....	194
19	APPENDIX 8 – SAMPLE DDEX MESSAGES	195
20	APPENDIX 9 – MOVIELABS COMMON METADATA SCHEMA (USED IN EIDR)	199

1 Executive summary

Linked Heritage (www.linkedheritage.eu) aims to enhance the quality, richness, re-use potential and uniqueness of the metadata contributed to Europeana (the flagship online cultural heritage portal of the European Union, accessed at www.europeana.eu).

One innovation compared with previous projects is the inclusion of commercial sector metadata in this remit, from four industry sectors corresponding to the four types of content found in Europeana: images, texts, sounds and videos.

Work Package 4 addressed this specific aim through Task 4.1, a detailed *Private Sector Business Exploration*, conducted through desk research, conversations with experts and practitioners in the four business areas, and analysis and comparison of relevant conceptual models.

The core findings were that each of the sectors discussed uses a range of metadata schemas and key identifiers that can be described as ‘best practice’ for that sector, and these standards are summarised in the table below. It shows roughly equivalent entities in each industry, plus libraries as a midway point between commercial and cultural heritage sectors, and the existing ATHENA infrastructure (used by the other work packages of Linked Heritage). Some of these identifiers and schemas are relatively new (e.g. EIDR) and hence are not yet in widespread use, whereas others are well established and used almost universally (e.g. ISBN). The images sector stands out as one where there is a lack of standardisation in current practice (though there are advanced proposals that may in future build a significant user base):

<i>Industry sector</i>	<i>Book publishing</i>	<i>Recorded music</i>	<i>Film and television</i>	<i>Images</i>	<i>Libraries</i>	<i>ATHENA</i>
Identifiers	ISBN ISBN-A ISTC	ISRC GRid	ISAN EIDR	PLUS ID	ISBN	ATHENA ID
Metadata schemas	ONIX for Books	DDEX	EIDR	IPTC PLUS XMP	MARC21	LIDO
Conceptual models	Indecs	Indecs	Indecs	–	FRBR(oo)	CIDOC-CRM
Examples of public-private partnership	Gallica 2 ENCLAVE	–	Ximon FindAnyFilm	–	–	–

Each of the commercial sector standards mentioned above is described in detail.

The report highlights a number of important contrasts between commercial and public-sector metadata. Chief among these are the commercial sector’s focus on metadata relating to *manifestations* (a class of resources, such as all identical copies of a book) rather than *items* (unique resources), the legal and commercial context that acknowledges that the metadata has significant

value, and the frequency, scope and commercial importance of continuous updating and ongoing management of the data.

Integration of data from these sectors will pose challenges in developing the semantic interoperability of the various metadata schemas, but the high degree of compatibility between the underlying conceptual models across three of the four commercial sectors, and the CIDOC-CRM model used in development of the ATHENA project, suggests that a semantic mapping exercise will be broadly workable.

This report highlights the advantages and challenges of using each type of data, and, with reference to the new Europeana Data Model and Data Exchange Agreement, the problems that must be addressed in future deliverables, both the technical (D4.2) and legal-commercial (D4.3).

There are a number of existing small- and mid-scale public–private initiatives which demonstrate the feasibility of cooperative ventures, but the limited life of some of these initiatives suggests that identifying a sustainable business model will prove critical if private sector metadata is to make a lasting contribution to Europeana.

2 Introduction

A world of exhaustive, reliable metadata would be a utopia. It's also a pipe-dream, founded on self-delusion, nerd hubris and hysterically inflated market opportunities.

Cory Doctorow, 2001

On two occasions I have been asked,—“Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?” ...I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

Charles Babbage, 1864

The initial quotation from Cory Doctorow represents the skepticism with which some view the attempt to create coherent, detailed metadata for describing (information) resources on the Web. Doctorow wrote that “observational metadata [based on statistical analysis of links to resources] is far more reliable than the stuff that human beings create for the purposes of having their documents found”. The self-contradiction in this objection is obvious, since clearly anyone could (even automatically) create many links to a resource they want to promote. In any case, major search engines are usually funded by advertising “that human beings create for the purposes of having their documents found”, and relying on precise the kind of controlled vocabularies and structured data for their results. As a recent collaboration between large search providers states, “Markup provides a consistent way for computers to understand the data on a page, and helps search engines display information usefully in search results.”¹ Charles Babbage’s amazement in 1864 at the thought that one could mechanically produce meaningful results without meaningful input is still relevant. Semantically rich, structured, expert-created data still works for search and discovery of information resources and still sells commercial products.

This report provides an introduction to the existing commercial sector best practice in content identification, description, supply chain and rights management metadata across four sectors within the European Union; thus presenting the essential results of Task 4.1, “Private Sector Business Exploration”.

The focus is on the description of the media industry and sectors within it, the technical specifications of standard identifiers and metadata schemas used within each sector, and the authoritatively recommended working practices in each. These descriptive and explanatory considerations will form the evidential, methodological and theoretical basis for the remaining Linked Heritage Work Package 4 (Public-Private Partnership) deliverables which will make recommendations for setting up partnerships between Europeana and the commercial sector.

This introduction explains the background to the work package and sites the report within it and the wider project. The research methods employed are then explained, followed by the results of the investigation, first in general terms then a sector-by-sector analysis in 4 parts:

- Books;
- Audiovisual;
- Sound recordings;
- Still images.

A summary and recommendations for the next Work Package 4 tasks and the wider Linked Heritage project complete this deliverable.

¹ See <http://www.schema.org/docs/faq.html#10>

2.1 Linked Heritage, Work Package 4

Aims of Linked Heritage

The Linked Heritage project's objectives are:

1. to contribute large quantities of new content to Europeana, from both the public and private sectors;
2. to demonstrate enhancement of quality of content, in terms of metadata richness, re-use potential and uniqueness;
3. to demonstrate enable improved search, retrieval and use of Europeana content.

Aims of Work Package 4

The work package as whole aims to enable public-private partnership in the contribution of content to Europeana, which until now has only been discussed as desirable and feasible in principle. Work Package 4, through this report, and future deliverables, lays the groundwork.

This work package has a central role in Linked Heritage and will contribute directly to all three objectives in the following ways:

1. Content;

Including commercial sector metadata will begin to address the problem of the “20th Century Black Hole”; the lack of contemporary materials, still covered by copyright and still of significant commercial value, which therefore are not immediately accessible to the cultural sector and its target audiences.² Before Linked Heritage there was no technical or legal framework for this.

2. Metadata enhancement;

The metadata created in the commercial sector is generally relatively rich and structured compared with, for example, Dublin Core or ESE. Inclusion of this data will constitute de facto enrichment of Europeana's collections.

Commercial metadata also often includes enrichments (“collateral content”) in the form of embedded previews of commercial cultural objects and/or Web links to such content. Types of additional content can include: book cover images, tables of contents, extracts or sample chapters; reviews and criticism; audio extracts; film trailers, low resolution pictures, etc.

Although this metadata is intended to enhance sales of products it is also “content” for Europeana in the sense that it enables a deeper appreciation of the cultural object under inspection before the user decides to access it. For example, a film trailer conveys impressions of images, sounds and a sense of the structure and plot of a film more directly and concisely than a set of key facts and a short written description might do.

3. Improved search;

Adding metadata records for items in the “21st Century Black Hole”, and connecting them with records from the cultural sector using a Linked Data approach, will begin to make visible relationships previously not explicit between contemporary and historical cultural objects.

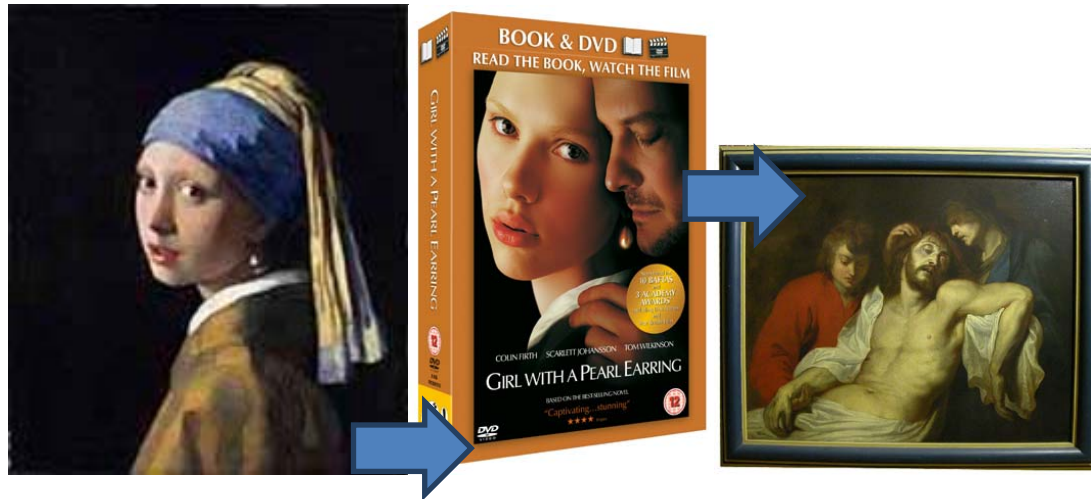
For example, a user might search for Vermeer's famous painting of the *Meisje met de Parel*³ and discover in the related content the 2003 film *Girl with the Pearl Earring*⁴. This might be available for

² See EC, 2011; “The New Renaissance”. Available from <http://dx.doi.org/10.2759/45571>

³ <http://europeana.eu/portal/record/92034/9FF847F41DB7B34F9482D876EE4EFFF093CB5F8D.html>

⁴ http://www.findanyfilm.com/Girl-With-a-Pearl-Earring-film_options~18722

purchase as a DVD or online download. Further, the metadata from this film might link it to Rubens' painting, *The Lamentation of Christ by Saints Mary and John*⁵, shown in the 2003 film hanging on Vermeer's wall (he was known to have owned the original); this fact appears via a search for Tracy Chevalier's novel on which the film was based⁶. It will be readily seen following the links in the footnotes to the previous page, that without the information from contemporary commercial connections, these historical heritage items would not be directly brought together.



Further, this deliverable feeds directly into WP4's later deliverables. Based on the best practice contained here, it will enable the specification of a recommended metadata format for commercial sector contributions to Europeana [D4.2]. It also describes the commercial and legal implications of reusing commercial sector metadata and hence gives the starting point for the legal framework and the business case to be constructed later in the project.

Finally, the best practice described in WP4 D4.1 explains how descriptive terminologies are used in commercial sector metadata, highlighting how they compare to EDM and how they might be expressed in SKOS / RDF. There are therefore essential concepts for consideration in WP2 (Linked data) and WP3 (Multilingual terminologies). The contacts and industry knowledge accumulated during this deliverable will form the basis of the network of industry experts which might assist in organising dissemination and creating training materials for staff of public and commercial organisations.

2.2 Background

Role of this Deliverable in the Project

This report is fundamental to the success of Work Package 4. Only by clearly identifying the most extensive and high-quality sources of commercial sector metadata and describing their technical and legal access points, can we proceed to make practical preparations for ingestion and develop the case for commercial partners to contribute. The first task of Work Package 4 according to the Description of Work is to "identify metadata models, rights models, aggregation and integration agreements, standards and specifications, which are in use in the private sector"⁷. This has been achieved and the

⁵ <http://europeana.eu/portal/record/09405a/AFDFDE168C29EBF9CDB0CE9D6A1B097EBAC4C5AA.html>

⁶ <http://www.amazon.co.uk/Girl-Pearl-Earring-Tracy-Chevalier/dp/0007232160>

⁷ Linked Heritage Competitiveness and Innovation Framework, Annex I – "Description of Work"

findings can be found in Section 4. Again, following the Description of Work, the research used the categories defined below to frame questions and present its findings:

1. Books (including ebooks, printed books and audiobooks);
2. Audiovisual (including filmed entertainment, television and film archives);
3. Pictures (including photography and still images).

Music and sound was added as a fourth category for three reasons:

- EDItEUR's understanding of the media industries suggests this is a sub-sector comparable to the others in terms of commercial importance, maturity of supply chain models and metadata standards and expertise;
- Europeana itself aggregates content in this area; it is one of four categories specified in the ESE metadata schema for all current Europeana aggregates⁸;
- the category is used to present search results in the main Europeana portal, and is listed as a distinct content category on the site itself:

"Ideas and inspiration can be found within the more than 15 million items on Europeana.

These objects include:

- Images – paintings, drawings, maps, photos and pictures of museum objects;
- Texts – books, newspapers, letters, diaries and archival papers;
- Sounds – music and spoken word from cylinders, tapes, discs and radio broadcasts;
- Videos – films, newsreels and TV broadcasts"⁹.

The work reported herein aims to give this foundation to the Linked Heritage project in terms of *content, metadata enhancement and improved search*, as described above, and also to create a basis for better understanding between commercial and public cultural sectors in the future, rooted in a shared understanding of both sectors' business practices and ways of conceptualising their work.

Inputs and dependencies

A conscious aim of this report was to build on the significant understanding of public sector best practice developed through *ATHENA* and highlight areas of similarity and contrast with commercial sector approaches. Thus reference is made to a number of *ATHENA*'s key deliverables¹⁰.

Since *this deliverable is breaking new ground for the Europeana community* there are no direct dependencies on other work packages; however, colleagues in all the other work packages were asked for information at relevant points in the task, to ensure an accurate presentation of the current state of progress thanks to *ATHENA*. The later deliverables in Work Package 4, in contrast, will necessarily draw on insights from the other Linked Heritage working groups.

2.3 Overview

2.3.1 The aim and originality of the research

This report is to lay the foundations for partnerships with the commercial sector to contribute metadata for in-copyright, in-commerce (paid-for) content to Europeana. For this reason the report begins with a

⁸ European Semantic Elements V3.4 at <http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd> ("type" element)

⁹ "Europeana – About Us", at <http://www.europeana.eu/portal/aboutus.html>

¹⁰ *ATHENA* Deliverables and Documents at <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>

discussion of the nature of the commercial sector content required by Europeana and a brief outline of the legal and technical framework that might be needed.

The originality of the research in Work Package 4 of Linked Heritage, going beyond previous and other current projects is:

- Enabling **contributions from all four sectors**, potentially under the same technical framework, so significantly more complex mappings and/or aggregations may be developed
- Much **larger scale contributions** than before (e.g. through EuropeanaConnect), requiring a substantially higher level of:
 - o Standardisation (e.g. maybe a LIDO-like mapping format);
 - o Best practice (*i.e.* data curation, rights management *etc.*);
- For the first time, **links to a retail environment** to enable convenient purchase of the items discovered (this is the FEP approach we are working with, described below:).

A clear and concise formulation of the fundamental business case for commercial partners to contribute product metadata to Europeana was given in the FEP (European Federation of Publishers) response to the EU Green Paper A-134-O (Vaisberg, 2010):

“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.”

2.3.2 Scope of the research and contents of the report

Existing examples of public-private partnerships, as well as commercial sector aggregation and integration standards, are discussed in the introduction to illustrate how large-scale contributions to Europeana are envisioned in practice.

There follows a concise explanation of the need for and nature of standards in the commercial publishing and media sector, outlining:

1. The typical media industry workflow and variations across sectors;
2. The reasons why identifiers and metadata schemas are used, especially with regards to rights management;
3. Some key differences between public and commercial sector metadata.

The four industry sectors covered in the report correspond to the four types of media discoverable through Europeana. They are:

1. Books (including e-books, printed books and audiobooks);
2. Audiovisual (including filmed entertainment, television and film archives);
3. Sound (including music and sound archives);
4. Pictures (including photography and still images).

For each sector the report details the current best practice in terms of:

1. Content, creative work and product identifiers;
2. Metadata schemas for content, creative works and products.

For each of the above frameworks for a given sector, essential technical and governance details are given. Finally, the report attempts an assessment of each framework from three viewpoints:

1. Degree to which the framework can be considered “standard”;
2. Level of market penetration;
3. Geographical adoption.

The conclusions of the report will evaluate the degree to which the above considerations present opportunities for building partnerships with commercial sector players to contribute content to Europeana, as well as obstacles to be overcome and recommendations for further work.

2.3.3 Limits of the project scope (what is not in this report)

In discussing identifiers and metadata standards, some might expect to find detailed discussions of:

1. Content standards (*i.e.* ways of formatting the intellectual content of a creative object);
2. Technical standards: ways of encoding and transmitting intellectual content digitally;
3. and potentially also physical formats of analogue cultural products.

This kind of discussion has been deliberately avoided since it is out of the scope of Linked Heritage WP4; except where necessary to clarify relevant concepts, or because content or technical standards normally include significant identifiers or metadata.

Similarly this report does not substantially address the use of identifiers or metadata in the cultural sector, except for comparison, as this was thoroughly addressed by ATHENA.

2.4 Links with other projects and work packages

Linked Heritage is a direct continuation of the ATHENA project¹¹ and its predecessors. It builds consciously on the networks, partnerships, best practice, standards and technologies developed within these, while aiming to make qualitative advances as described earlier.

Linked Heritage Work Package 4 is not unique in aiming at enabling ingestion of cultural industries' metadata (though see section 2.3.1, "Originality..." above); the Europeana Connect project¹² formed commercial sector partnerships, created audio recording metadata mappings and ingested significant commercial content in the audio domain via the DISMARC aggregator¹³.

Other projects which have built aggregators which might ingest commercial sector content (or provide a working model for this goal) are Europeana Libraries¹⁴ (in the books domain) and European Film Gateway¹⁵ and EU Screen¹⁶ (for AV). Linked Heritage will draw on best practice and lessons learned from all three.

EDItEUR, with MVB, ICUU and mEDRA, is currently a partner on the ARROW+ project¹⁷ to build and promote books-in-print databases and implement the orphan works rightsholder search infrastructure elaborated by the first ARROW project¹⁸. Linked Heritage has and continues to benefit from its expertise in the book publishing domain and copyright field, as well as of the networks build in ARROW Plus also in the image domain..

Other immediately relevant work packages within Linked Heritage are:

- *WP2 – Linking Cultural Heritage Information*

¹¹ See <http://www.athenaeurope.org/>

¹² See <http://www.europeanaconnect.eu/>

¹³ See <http://www.dismarc.org/>

¹⁴ See <http://www.europeana-libraries.eu>

¹⁵ See <http://www.efgproject.eu/>

¹⁶ See <http://www.euscreen.eu/>

¹⁷ See <http://www.arrow-net.eu/>

¹⁸ Full reports of the first ARROW project: <http://www.arrow-net.eu/resources/arrow-project-public-reports-deliverables.html>

This work package assisted with understanding the Linked Open Data (LOD) context, the LIDO metadata harvesting format, and the CIDOC-CRM ontology. The working group included contacts with useful expertise in photograph metadata;

- *WP3 – Terminology*

Work Package 3's group assisted with SKOS and other terminology-related queries, as well as contacts in Belgium and France;

- *WP 5 – Technical Integration*

This group provided advice on technical solutions for aggregation and automatic metadata mapping.

3 Europeana Data Model and Commercial Content

As mentioned above, although other projects and aggregators have made isolated attempts and progress in analogous areas, Linked Heritage aims to do something fundamentally new in enabling large-scale contribution of rich commercial metadata across all four content types.

The value proposition to the commercial contributor, which will be developed in detail in the third and final deliverable of this Work Package, is to offer a new discovery platform for commercial products, with an inbuilt sales channel, and association with great works of artistic, historical and cultural value.

In order to demonstrate the value of commercial sector metadata to Europeana it is important to consider the existing context in which contributions are made to Europeana, the metadata schemas currently supported, and the terminology used to describe the mapping and ingestion of contributions to Europeana. In particular we wish to clarify the two distinctions:

- Europeana Content versus commercial sector metadata;
 - o Between storing the actual substantial cultural object (“content”) and...
 - o Aggregating merely a representation of that object, a “surrogate record”¹⁹ composed of:
 - metadata [numerical and textual]
 - a preview [textual, visual or audio]
- Previews in Europeana versus “marketing collateral” used in cultural industries;
 - o between metadata and previews in the commonly used sense above and...
 - o the broader concept used in the cultural industries

These distinctions are keys for understanding why Linked Heritage aims to enable commercial sector partnerships, the detailed commercial metadata landscape at the time of writing, and the technical and legal-commercial issues that will be faced in subsequent tasks.

3.1 The kind of content required by Europeana

Content versus metadata: aggregation

Europeana provides access to cultural heritage indirectly. Instead of aggregating content (digital texts, videos, audio files and photographs) it only aggregates metadata from local providers of content in EU countries²⁰. Sometimes this is indirect, via intermediate national or sectoral (e.g. only for audio archives) aggregators. Users discover information about content, including in many cases image previews, and navigate among such surrogates records or aggregates, and, if they wish, follow links to content providers' websites where the full content (or a higher resolution digital representation of the object) can be found.

EDM aggregates

These concepts of access to *content* via *links* discovered in *aggregated (resource) metadata* are used in the Europeana Data Model (EDM) to more specifically define the type of metadata Europeana contains. In fact EDM describes entities called “aggregates” which embody the nexus of the outward links to content providers, and inward links to description and previews stored in Europeana. This is

¹⁹ See Taylor, A. G. (2004).

²⁰ See Dekkers, *et al.*, (2010).

the concept of aggregation as specified by Open Archives Initiative²¹. The papers *The Europeana Data Model (EDM)* by Doerr *et al.* (2010) and the *EDM Primer* (Isaac, A. *et al.*, 2010) elaborate this concept more fully.

EDM content types – content form and media

There are four main media²² types admitted by EDM: texts, images, videos and sounds. A fifth 3D model category is in development. The EDM media types cross-cut the content forms that may be available in the commercial sector, as explained below.

3.2 The rich metadata found in the commercial sector

In distinguishing typical Europeana contributions from the possibilities offered by the commercial sector, the potential for semantic enrichment of Europeana will also become clearer.

Previews versus Marketing collateral

In Europeana's terms, what content providers call "metadata" becomes "Europeana content". In all sectors, the digital supply chain metadata includes previews of the core products²³ offered in the form of "marketing collateral"²⁴ used to differentiate a given product from similar offerings at several stages of supply chain. "Marketing collateral" will be the term used here; the same concept and practice exist in all sectors, whatever term is used. This in effect is a form of creative content in miniature – extracts, previews, reviews, abstracts and derivative works of many kinds:

Sector	Books	AV	Music	Images
Marketing collateral	Synopses	Synopses	Album cover art	Thumbnail images
	Sample pages or chapters	Trailers and Clips	Promotional videos	Photographer profiles
	Author profiles	Reviews	Audio clips	Listings of related images
	Reviews	DVD extras	Reviews	
	Book cover images	DVD cover images	Lyrics	
	Tables of contents	Cast / crew profiles	Band biographies	
	Subject/genre coding	Contents lists (scenes, chapters, extras)	Track listing, timings <i>etc.</i>	
	Full text of books (for search only)			

There is a wide variety of practice, ranging from describing and linking to Web resources, making collateral files available at specified locations for download, right through to including collateral content within the metadata file itself. It is of course technically possible to simply offer entire packages of core

²¹ For the full details of OAI aggregation, see <http://www.openarchives.org/ore/1.0/datamodel> and <http://www.openarchives.org/OAI/openarchivesprotocol.html> for the relevant standards' specifications

²² In Indecs, media are known as "genres" but this term is not commonly understood thus. See Appendix 1 – Glossary of Terms for clarification

²³ See Indecs, 2000, p. 36: "...metadata may equally be comprised of non-textual creation types, such as diagrams, thumbnail images, audiovisual and audio clips, watermarks and so on..."

²⁴ See EDItEUR, 2009a, p. 5.

plus collateral content to all downstream partners, with a metadata wrapper describing the nature and intended uses of each file. Due to the logistical factors of cost, speed and capacity to transfer large data packages, this is probably not the norm for any sector (except perhaps images where there is little “collateral”, and the individual core content file is already pre-packaged in a metadata wrapper by the camera, and in e-books in certain proprietary formats).

3.3 Semantic Enrichment Benefits from Commercial Metadata

The main potential benefit Europeana's users is richness and variety of the collateral content which may be available, much of which will already be provided in the appropriate language for a given EU state for marketing purposes.

One example of a hypothetical, semantically-enhanced search has been cited in the introduction to this report. In general, commercial sector metadata is suitable for linking and creation of relationships between commercial products, their abstract creative content, and creators because of

1. its orientation towards marketing and sales, downstream in the supply chain;
2. its intrinsic nature supporting the upstream value chain, principally sales reporting, rights management and royalties.

Point 1 refers to the fact that enriching the information about commercial content with details about its creator, its distinctiveness, quality and critical reception are in the direct interests of the producers and distributors who rely on customers' choice and purchase of their products for income. Although the motivation is clearly commercial and financial, it is one that is weighted in the interest of the recipient of such data, in that consumers are not obliged to purchase content, and on the contrary, expect real value from their limited time and budget for cultural purchases.

The second point highlights the enrichment of metadata in order to make it usable within the supply chain itself, allowing greater efficiency in selecting and distributing content at different points from creation to sale. While this enrichment may aim at simplifying processes with the industry (for example, bringing together different versions of cultural work to ensure royalties are correctly paid to the creator) it has value for researchers, students, discerning purchasers and the interested citizen beyond the intended use. The necessary standardisation of names, descriptors and identifying codes for use along the supply chain makes semantically robust linking feasible.

Both of these aspects of commercial metadata enrichment affect the structure of identifiers and metadata schemas developed to manage them, and make them useful for the project of linking cultural heritage in a Semantic Web context.

In all sectors there is a strong drive to make identifiers of products unique, persistent and interoperable with other registers of works, interested parties and indeed related products and works so that rights, royalties, sales, and workflows can be tracked and optimized. This means that creating links between metadata records, identifying ontological associations, and mapping aggregated records to the original provider are all facilitated. Identification is more or less established in different sectors, but the problem of increasing granularity²⁵ of digital content and complexity of digital products – and therefore the business needs and working principles – are the same across all sectors.

This increasing granularity and complexity has the consequence that metadata schemas tend towards specification of detailed vocabularies and hierarchies of entities and their attributes so that the correct product can be addressed to the appropriate market, and rights and royalties correctly managed. This means that detailed descriptive information is available and new semantic mappings will become possible.

²⁵ See <http://www.bic.org.uk/files/pdfs/uniquid.pdf> – section on “Granularity”

3.4 Limitations of Europeana top-level Aggregation

Having stated that an immensely rich discovery experience is potentially deliverable from commercial sector partners, it is important to frame this possibility within the actual constraints imposed by the top-level aggregation model of Europeana. While detailed technical and legal arrangements will not be recommended here, it is incumbent on this report to make brief mention of the most obvious potential difficulties and point out positive directions that might be taken.

New Data Exchange Agreement (DEA)

The new Europeana Data Exchange Agreement²⁶, due for final implementation from the first day of 2012, stipulates that all Europeana content providers will agree to Europeana's re-publishing the data they have contributed as Linked Open Data under the Creative Commons CC0 licence²⁷ allowing any re-use, including for commercial purposes and without attribution, of that contributed data. This follows on from a Linked Open Data pilot²⁸ and development of a range of API services²⁹. It is important to note that the DEA only applies to the textual metadata (not image previews) held in Europeana (other data held by the content provider and not given to Europeana is unaffected). The CC0 license entails waiving all legal rights in the data provided, which may deter potential commercial organisations from contributing anything for the following reasons:

- Loss of control of the data (CC0 is irreversible and in the most strict national legal framework would default to a public license granting unlimited rights);
 - o Lack of guarantees for the context in which data will appear (in the absence of a specific contract or terms)
 - o Loss of reputation due to unfavourable, fraudulent or potentially defamatory reuse
 - o Loss of attribution and recognition of work and resources invested
- Challenging the business models of existing aggregators of commercial data who generate income by collecting, enhancing and building services around corpora of this data;
- Cutting across existing API licensing and partnership agreements in commercial and public sector contexts.

The (commercial) re-use of textual metadata elements will be allowed and encouraged. Other intellectual property stored in the top-level Europeana portal (principally the image previews of book/CD/DVD covers, or lo-res versions of pictures), is clearly defined currently³⁰: it is "assumed to carry the same rights as the underlying object". However, this should be more explicitly displayed at the search results level, especially as re-use of data is actively encouraged.

Given that rights are so much at the core of digital content in the cultural industries (see Rust, G., 1998³¹) a more nuanced description of rights may need to be developed in order to meet the expectations and requirements of the commercial sector. The contractual rights of media organisations are often delineated by time, territory *etc.*, and as the granularity of digital products increases, so do the rights in them become more granular and diverse.

Basis in Dublin Core

²⁶ See <http://version1.europeana.eu/web/europeana-project/newagreement/>

²⁷ About CC0 – "No rights reserved": <http://creativecommons.org/about/cc0>

²⁸ See <http://version1.europeana.eu/web/lod/home>

²⁹ See <http://version1.europeana.eu/web/api>

³⁰ See presentation on CC0 licence:

http://version1.europeana.eu/c/document_library/get_file?uuid=c89e37f8-6d75-4375-80aa-1d1faf0ce693&groupId=10602

³¹ Available at <http://www.dlib.org/dlib/july98/rust/07rust.html>

Another limitation of the Europeana portal is the conceptual ambiguities and semantic overlaps of the Dublin Core descriptive elements³² upon which ESE and EDM are based. Dublin Core is also a “flat” schema. Both of these limitations mean that rich commercial sector metadata will lose much of its richness, and often gain some ambiguity, in mapping to EDM/ESE. While it is true that ESE and especially EDM add more elements and relators to the Dublin Core elements, because these form the semantic core of both schemas, the fundamental, well-acknowledged problem is not solved by adding qualifiers or more terms.

The potential loss of detail and precision in moving from an event-based metadata model (such as is found in the commercial sector) to a resource based model like Dublin Core is well documented. Its effects can be seen in the reduction of elements to a common denominator like coverage in the following example:

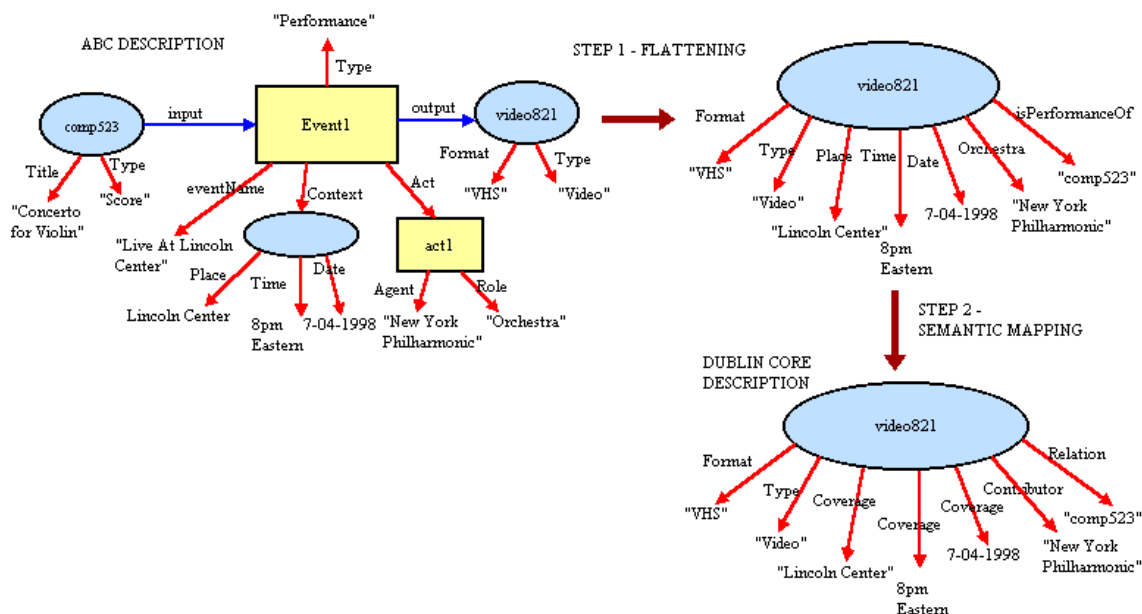


Diagram of mapping to Dublin Core from Lagoze, Hunter and Brickley, 2000.

Basic Dublin Core³³ does not take into account essential, non-negotiable aspects of commercial metadata, both in its logical structure and in its inclusions, omissions and choice of elements. In a foundational paper by Rust, 1998, these four principal difficulties are identified:

- “First, as a core data set, it is compromised [...] There are significant question-marks over the majority of the fifteen elements. Their grouping into three overall categories (*Content*, *Intellectual Property*, and *Instantiation*) doesn't really help...;
- Secondly, its attributes, terminology, and examples are derived principally from text-based creations [...] and make it likely to appear largely irrelevant to owners and publishers of audio, audiovisual, and abstract works, despite the fact that a cross-sector vocabulary is now essential;
- Thirdly, 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;

³² Park and Childress, 2009.

³³ See Dublic Core Metadata Element Set, V1.1 <http://dublincore.org/documents/dces/>

- Finally – and this is the main difficulty with Dublin Core for rights owners – it views rights metadata as an extra (15th) element or set of elements, not recognising that, in fact, it embraces 13 of the other 14 elements...”.

As discussed in the previous section, the rights aspect of commercial metadata is essential to their character and is not a secondary question – it even affects the structure and semantics of the raw data involved, as well as their interoperability and discovery characteristics.

Suggested enhancements to EDM are envisaged as part of Linked Heritage (work packages 2 and 5), including enriching the semantic possibilities of elements through linked data approaches³⁴, more specific definition of terms and mapping to existing vocabularies. Simple addition of essential elements from the commercial sector, new types of object covering audio, AV and image attributes and more rights-oriented structure would significantly improve the model.

³⁴ See Linked Heritage deliverable D2.1 and the following documents on Linked (Open) Data:
<http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/> – W3C Library Linked Data Incubator Group Final Report
<http://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-20111025/> – W3C Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets
<http://www.w3.org/standards/semanticweb/> – W3C, Semantic Web homepage
<http://lod-lam.net/summit/about/> – International Linked Open Data in Libraries, Archives, and Museums Summit (“LOD-LAM”)
<http://www.clir.org/pubs/reports/pub152/Stanford%20Linked%20Data%20Workshop%20Report%20FINAL%20111024.htm> – recent (Oct 2011) report on library Linked Data

4 Research Methods and Approach

To prepare this report, a mixed methodology was employed, consisting of networking and conversations with interested experts within the UK and across relevant organisations EU-wide, and desk research using EDItEUR's substantial in-house expertise, the author's professional association's resources³⁵ and the Internet.

Some of the results of this research have already appeared as the background and introduction to the report, as they are essential preliminaries to the core findings. This section on methods will explain how these concepts and facts were discovered, digested and harmonised.

4.1 Industry contacts and best practice network

Contacts in commercial cultural content industries

The primary approach of this research was interviewing industry sector experts found through EDItEUR's network of contacts in similar standards governance, trade associations and metadata intermediary roles, and the wider augmented network enabled by the other partners in Work Group 4. The delegated nature of the commercial sector's interconnections comes into play here, to enable a comprehensive overview view without directly surveying all the players in any given field.

Since standards and intermediaries exist specifically to remove the need for all institutions to maintain their own experts and registration systems, the delegated experts act as nodes of connection and custodians of common understanding between at least the main commercial players and the main representative bodies for large, medium and small companies.

4.2 Desk Research

In addition to interviews with industry experts the following methods were used to develop expertise:

Web searches

Any documentation that was available on the open Web or with free registration (a common practice for organisations publishing standards) was obtained through these means. Documents used for the report are referenced at the end of this report, and securely archived locally at EDItEUR in case of changes or removal from the Web.

Literature searches

In addition, literature searches for a small number of core key words (e.g. names of all the identifiers and metadata schemas covered here, "commercial sector", metadata, "standard identifiers" etc.) were performed within Google Scholar, the Proquest Library Science database and the full indexes of the following journals (the set of titles available through CILIP membership, as noted above):

- Journal of Information Science (JIS);
- Journal of Librarianship and Information Science (JOLIS);
- Health Informatics Journal;
- IFLA Journal.

³⁵ Chartered Institute of Librarians and Information Professionals (CILIP). See <http://www.cilip.org.uk/>

Research Instruments

During the course of the research described above two main research instruments were developed:

1. A theoretical basis formed from conceptual frameworks in the public heritage and commercial cultural industries sector – this is embodied in the next section (“conceptual frameworks”) and the Glossary of Terms in Appendix 1 (acronyms are listed in Appendix 2). This glossary forms a network of links between these two domains, and includes terms not used in the main body of the report, but which link between other terms to add context and richer explanation;
2. A format for describing standard identifiers and metadata schemas – this is described in the Industry Sector Findings section, under the heading 8.1 Presentation of results. It draws on formats used in previous research in both commercial and cultural heritage sectors, and includes substantial explanations of key technical terms and concepts.

These research instruments therefore are not merely convenient structures, but represent significant knowledge and expertise in themselves. They can be used to inform further research in these areas, as well as practical applications of the data they operate on.

4.3 Conceptual frameworks

The terminology used to describe and discuss metadata varies considerably even within one field of activity. Even the term “metadata” has been thoroughly debated. Because this report addresses four distinct sectors of creative industries, and because even these fields are not all equally coherent and unitary, the problem of varying terminology is made more difficult and defining a common set of terms will be a valuable contribution in itself.

Why conceptual frameworks?

To address this problem, and others, metadata practitioners develop conceptual frameworks applicable to large domains (e.g. Indecs for e-commerce, FRBR for library cataloguing, CIDOC-CRM for museum objects, discussed below). Having modelled the underlying entities, relationships and processes generally involved in the domain, it is possible to make metadata schemas more coherent, and ensure that identifiers are pointing to the right kind of entity.

The relevance of these models to Linked Heritage is that by understanding and comparing the viewpoints of each domain and sector we can better understand what each metadata and identifier standard aims to do, and find common language to present our aims and requirements. The concepts here were used to inform the harmonisation of terminology applied to the findings of the report during the research, thus forming both part of the method and results.

Choice of domains and models

Mentioning <Indecs> as representative of commercial sector metadata is an obvious choice, especially as it explicitly informs many of the main standards in this sector (see books and music sections below).

The use of CIDOC is also clear due to its link with the LIDO standard used for ATHENA and Linked Heritage metadata aggregation.

FRBR (and FRBRoo) was chosen as a representative of the library (and library-like) sector’s view, which collects and curates ‘less unique’ published literature (and also other published, mass-produced works like audio and audiovisual recordings and photographs) but also more unique objects (like rare

books, original phonograms and photo negatives)³⁶. Thus the library domain acts as a half-way house in the “uniqueness spectrum” between “pure” cultural heritage institutions and the commercial cultural industries. The spectrum, although in reality a continuum, to some degree inhabited across the whole range by both metadata creators and conceptual models, can be described as below to highlight the areas of main interest:

Uniqueness spectrum			
Entities described	Unique objects	Specific (sometimes unique) instances of more generic products	Classes of identical products and their creative content
Metadata creators	Museums Galleries Archives	Libraries (of very many different media and purposes)	Cultural industries
Conceptual models for metadata	CIDOC-CRM	FRBR(oo)	Indecs

The library sector also routinely collects and documents large volumes of commercial sector products, and the FRBR(oo) model provides halfway point to illustrate the continuity. Even more importantly, it demonstrates in a concrete way that harmonisation of the CIDOC-CRM and Indecs viewpoints is possible.

³⁶ See, for example, CIDOC’s FRBR tutorial at http://www.cidoc-crm.org/docs/2008-01-25_FRBR_tutorial.ppt for this comparison

5 Commercial Sector Metadata Landscape

The material presented in this section is applicable across the entire commercial sector. It draws together top-level findings and describes themes and contrasts found in comparing the subsectors within commercial cultural endeavour, to prepare the reader to understand the detailed sub-sector analyses in the coming section, Chapter 6.

5.1 European Cultural Industries

The first question in describing the cultural industries in the EU is: which industries? The answer enables us to define the objects of this study and draw comparisons between them.

5.1.1 Size and Scope of the Cultural and Creative Economy in the EU

Attempts to define cultural industries have produced definitions that are broadly similar but differing greatly in detail. Even the attempt of Potts *et al.* (2008), while dismissing previous definitions “according to industrial sectors” and “the type of industrial activity in which a firm is engaged and the nature of its material inputs and outputs”, in favour of a “social networks” approach, ended up with a model which includes:

- advertising, architecture, media, ICT software, *etc.*
- film, TV, music, fashion, design, *etc.*

These are the same types of core industries and products envisaged by all of the main reports and guidelines extant. A summary of these convergent versions was made in *The Economy of Culture in Europe* (KEA European Affairs, 2006, for the European Commission). KEA European Affairs (2006) described³⁷ the European cultural and creative industries at the time as a significant contributor to the EU’s GDP, economic growth and employment. It compared favourably to production, manufacturing and processing industries in terms of GDP at 2.6% of the total for Europe; it also showed a 12.3% higher growth in 1999–2003 than the general economy. Recent economic crises must have had an impact on its health, but as the KEA report points out, the statistics probably under-represent this group of industries. KEA identified “sectors” and “circles” of related economic activity, categorised as overleaf:

³⁷ See the full report at <http://www.keanet.eu/ecoculture/studynew.pdf>

	“Circles”	Sectors / subsectors	Characteristics
“Cultural sector”	Core arts fields	Visual arts (including Photography) Performing arts Heritage	Non industrial activities Not systematically copyrighted
	Cultural Industries	Film and video Television and radio Video games Music Books and press	Industrial activities aimed at massive reproduction Outputs are based on copyright

Table adapted from KEA European Affairs (2006) p. 3.

5.1.2 Correct scope of “cultural” or “content” industries

Although the KEA (2006) analysis is helpful for defining an industry at the level of Europe it has some shortcomings, in that it includes photography in the Core Arts Fields which it characterises as “potentially copyrighted works... but... not systematically copyrighted”. However, photographs are equally entitled to copyright protection under existing legal frameworks.³⁸ The current report will consider photography as belonging firmly within the same “cultural industries” category as the other three areas of interest.

Interestingly for the purposes of this report, KEA placed the remaining three of the four industry sectors covered by this report in the same “Circle”, the “Cultural Industries”. The exclusion of photography *per se* reflects its status as a disparate community of artistic producers (sometimes) selling their works through a network of businesses that normally sell in bulk to other businesses within the cultural industries circle. This exceptional nature of images sector will become apparent in the rest of the report. In this report the correct definition of the commercial content industries follows Rust (2005): “**the major content sectors of text, music, images and audiovisual**”.

5.1.3 Cultural Industry sectors in detail

Because the cultural industries are based on copyright to protect and define the intellectual property vested in the “content” they produce, manage, license and sell, and because they aim at mass reproduction of this content, it is possible to draw common outlines and comparisons between reasonably similar sectors.

All of the cultural industries are further characterised by the existence of a supply chain which normally includes a minimum of four steps: creation, publication, distribution and retail. The specific definition of these stages can be argued to expand them into more, for example, “registration” of a creation may precede “publication”, or “retail” may be broken down into “marketing”, “consumption” etc.

³⁸ See Berne Convention, Article 7:

http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html#P127_22000

However, the important concept to grasp is the more-or-less linear chain of processes and actors intermediating between the creator and consumer of commercial creative products. In fact, it might be argued that this intermediation is what makes the cultural industries commercial, as normally some money or value is exchanged at each step (see below for more on “value” versus “supply chains”).

This is summarised in the Indecs “model of commerce”: *People make stuff. People use stuff. People do deals about stuff.*³⁹

The commercial view of the entities produced by cultural industries, formalised by Indecs is a kind of ontology (see Ding, 2001) which has informed some, and helps make sense of all identifiers and metadata used for commercial products. The relevant ontologies for the library and heritage sectors are, respectively, FRBR⁴⁰ and CIDOC-CRM⁴¹. It is instructive to note that Indecs is event-based, and neutral as to commercial model, so can include licenses and transactions relevant to publicly-funded sectors as well: transactions at zero financial cost can still be modelled as transactions. Conceptual models were discussed in depth in section 4, again highlighting that the commercial and cultural heritage sectors represent different viewpoints within a continuum rather than completely disjoint areas of activity and interest.

5.1.4 Modes of production and classification of products

The four sectors are differentiated roughly according to the modes of original production and of consumption; each sector's products are consumed using different combinations of senses and thus require specialised techniques of production, identification and description:

Industry sector	Sensory aspects*	Secondary content
Books	Visual (text) [audio – audiobooks only]	Finished product normally includes images, at least e.g. for cover, illustrations
AV	Visual (text + images) Audio	Textual content (excepting scripts) is normally insignificant by quantity but may be integral to the content
Music	Audio	Finished product normally includes images, at least e.g. for cover, description of content, lyrics etc.
Photography	Visual	Normally minimal

* primary product content only

It is worth noting at this point that the above classification by sensory mode restricts certain artistic works to one expression and media type (although not at all by carrier type) – for example, a musical work for our purposes will generally be a recorded performance, not a written score, or digital file containing musical notation⁴². This is partly due to the fact that cultural industries ultimately produce and sell fixed objects manifesting their works in a practical form for delivery, and partly due to the

³⁹ <indecs> Principles, model and data dictionary, page 4.:

http://www.doi.org/topics/indecs/indecs_framework_2000.pdf

⁴⁰ FRBR Final report: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

⁴¹ CIDOC-CRM official version: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf

⁴² See ISBD Area 0 for elaboration of this concept: http://www.ifla.org/files/cataloguing/isbd/area-0_2009.pdf

constraints of the EDM media types. For the purposes of Linked Heritage this represents a convenient and very practical simplification. Currently, Europeana distinguishes the following media types and content carriers⁴³ (potential commercial items for inclusion in Europeana are **bold and underlined**):

- Images – paintings, drawings, maps, **photos** and pictures of museum objects;
- Texts – **books**, newspapers, letters, diaries and archival papers;
- Sounds – **music** and **spoken word** from cylinders, tapes, discs and radio broadcasts;
- Videos – **films**, newsreels and **TV broadcasts**.

Note that “spoken word” sound recordings will appear in the “books” category because they are classed as “audiobooks” by the book publishing industry. Again, it is apparent that the mapping from heritage sector to cultural industries will not always be one-to-one.

Though archival versions of historical newspapers are available, and Europeana could potentially get access to their online content through some sort of deal (as they are often commercial products available on subscription), they were excluded from Linked Heritage’s scope.

The supply chain(s) present in each industry use metadata and identifiers to maximise the chains’ efficiency, in terms of working practices, but also in the design of the IT systems used internally by companies.

Commercial IDs and metadata are first and foremost practical tools, and historically were not based on any comprehensive theoretical framework, although as mentioned above, digitisation of content and supply chains has driven theoretical work that has been used to inform more recent initiatives such as DOI and ONIX⁴⁴, EIDR⁴⁵ and DDEX⁴⁶. Hence in each sector and sub-sector different approaches to processes, data and systems have resulted in differing standards; there is no one “master system” uniting all cultural industry metadata and products.

Media and publishing industries are known to exhibit high levels of integration (see for example, Albarran, 2010) across the four distinct sectors identified by Linked Heritage and used to structure this report. Digital creation, online distribution and e-commerce have tended to intensify this phenomenon as commercial books, AV, images and music are increasingly regarded as sub-categories of digital “content”.

This is not to say that whole sectors have merged, but that some elements of their supply chains now share technological services. As a result, metadata standards, best practices, supply chain business needs and intermediaries’ solutions may change significantly. Some sub-sectors, “can increasingly be seen as the different industries that they are (previously united by a single physical product)” (Bide, 2011a). Signs of these developments are highlighted throughout the report, as are relevant legal and commercial aspects of the actual metadata standards, frameworks and practice, in order to both realistically contextualise the descriptions, and to anticipate the work ahead.

⁴³ See <http://www.europeana.eu/portal/aboutus.html>

⁴⁴ “DOI is an implementation of the indecs metadata framework, which was also a guiding principle of the ONIX development”. See: <http://www.doi.org/news/020319-Editeur.html>

⁴⁵ See EIDR, 2011b. Because EIDR implements DOI, it inherits DOI’s foundation in the <indecs> model. See http://www.doi.org/about_the_doi.html for more on DOI (Digital Object Identifier)

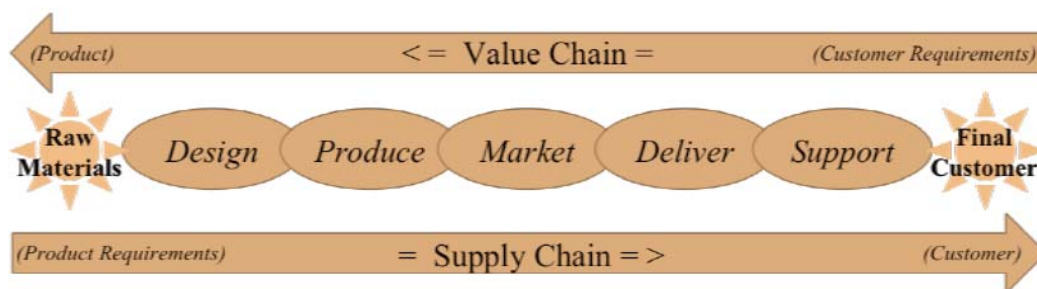
⁴⁶ See Bide, M., 2011. <http://dx.doi.org/10.3998/3336451.0014.108>

5.2 Supply chains

In order to understand the detailed sector descriptions later in this report, the discussion that follows will give the reader a basic understanding of the supply chain concept as applied to cultural industries and the origins and nature of the metadata frameworks found there.

5.2.1 Supply chains, identifiers and metadata schemas

The concept of a supply chain has been used to describe each industry sector in this report, for the simple reason that this viewpoint best explains the need to create standard identifiers and metadata schemas. Researchers in media industries (see e.g. Chrissey, 2010; Bloore, 2009) have discussed the merits of using concepts such as “value chains”, “value systems” and “supply chains”. It appears that these can be synthesised in a diagram like that below, showing how in a generalized market situation the production flows from the industry to the customer, while value returns the other way to recoup the costs of production (and in most cases make a profit):



Value Chain vs. Supply Chain comparison (Feller et al., 2006; quoted in Crissey, 2010).

As noted above in section 3.3, commercial sector metadata has important uses in both directions. Supply of the product to the customer is made more efficient (through communication of operational metadata between supply chain partners) and effective in terms of sales results (through marketing metadata). The operation of the value chain is improved by sales reporting (which communicates “customer requirements” back to producers in terms of actual demand) and rights and royalty management (enabling sustainable creation of existing and new products).

The simplified media industry supply chain below summarises the creation, production, distribution and exhibition (or consumption) of media content. Albarran (2010, p. 57) describes an idealised process of media value creation leading from the creator to the consumer through four stages:

- Content creation – the “artistic work” is conceived by an author, songwriter, photographer or director;
- Production – the work is fixed in some sort of final form; a final book draft, mix of a song, raw image, final cut of a cinema film;
- Distribution – through several sub-processes, the final version of the production is made available on a local level through a network of manufacturers and suppliers; book printers, CD manufacturers, photographic developers and printers, film duplicators. In all cases, these can now take the form of electronic distribution, more or less integrated into production; of course the supply chain ends with some kind of wholesalers (or file aggregators) and retailers (online or physical);
- Exhibition – also known as consumption of the media product.

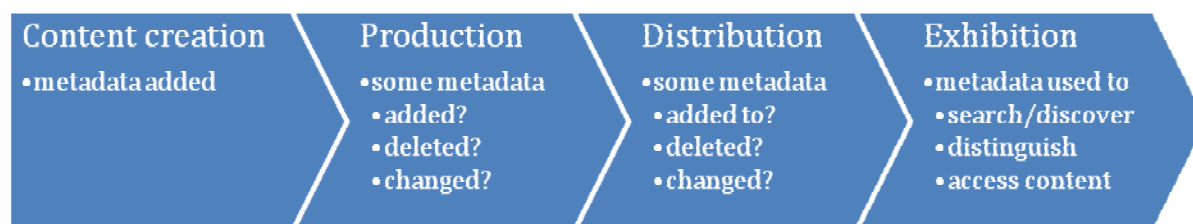


Diagram adapted from Albarran, 2010; MWG 2010

In this report “supply chain” will generally refer to all the production and distribution processes between creation and exhibition to distinguish this central (and often invisible) portion of the process, and the intermediaries who operate there from the more obvious creation and consumption of media.

5.2.2 Types of cultural industry businesses

The kinds of institutions in the public cultural heritage sector may be well known to most Linked Heritage partners. These institutions will include universities and government ministries, the “cultural sector” itself: galleries, libraries, archives, museums (GLAM), and a range of other intermediating and ancillary organisations and institutions, many of which remain in the publicly-funded sphere, but some of which will be for- or not-for-profit bodies technically within the commercial sector.

It is important to bear such crossover companies in mind when considering the business types typical in the commercial media sectors, as in some cases they deal with public sector *and* commercial clients; their business processes span the divide and can be instructive for our purposes.

The following table attempts to show that roughly analogous organisations exist at a given stage within each industry’s supply chain – it must be stressed that this is a substantial simplification of the real, extremely complex business. Parties and organisations mentioned are illustrative examples and not exhaustive.

Type of organisation	Sub-Sector			
	Books	AV	Music	Images
Content creators	Authors	Script writers Directors Actors	Lyricist Composer Recording artists	Photographers Illustrators
Collective (rights) Management Organisations (CMOs)	Reproduction Rights Organisations	Retransmission Rights Organisations	Song rights agencies Royalties distribution agencies	Visual artists’ representation bodies
Producers	Book publishers	Studios Production companies	Record producers	Photo agencies Studios

Type of organisation	Sub-Sector			
	Books	AV	Music	Images
Distributors	Book distributors	Film distributors Digital cinema distributors	Digital music file aggregators CD distributors	Photo agencies Photo libraries
Supply chain Intermediaries	Books-in-print databases Metadata aggregators / vendors	Film databases TV listings	Album information services	PLUS registry
Exhibitors	Bookshops (online and 'bricks and mortar') Libraries	Video On Demand (VOD) DVD retailers Cinema operators	Music download sites Bricks and mortar retailers	Print retailers [picture researchers in other media firms]

5.3 Commercial sector distinctives

Commercial metadata can appear similar to the metadata used in the cultural heritage sector. To some extent this is true, since both sectors need to:

- Identify the entities they deal with ("resources", "releases", "products" etc.);
- Describe entities;
- Relate resources to e.g. each other, their creators, access locations and methods.

On one hand there is a great deal of overlap in the focus on describing the entity. On the other hand, the focus of commercial metadata is weighted towards how the entity is produced, how it and its associated rights change owners, *i.e.* price, availability, order point, *etc.*

5.3.1 What is metadata?

A typical definition of metadata in the cultural heritage sector is "data about data", for example:

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information."⁴⁷

This definition is certainly useful for a general introduction but much more specific and operationalised terms are needed to cross the boundaries of the public and commercial sectors. This report will use the definition provided by the Indecs (2000) framework:

"An item of metadata is a **relationship** that **someone claims** to exist between **two entities**."

The added bold terms in the quotation above highlight the key factors involved. The framework whereby two entities (whether mere data or some physical object, person, event *etc.*) are linked by a relating term brings to mind the RDF triple syntax used in Linked Data. The additional commercial

⁴⁷ See NISO, 2004.

aspect of this definition is the fact that the metadata in question here represents someone's claim about the relationship, e.g. of ownership of commercially valuable intellectual property rights in a piece of data representing creative work. The extra provenance relationship implied here, incidentally, goes beyond the simple RDF triple model⁴⁸.

The basic problem of metadata standards, modelling, mapping and identification is found here and has two entirely interlinked aspects:

1. The provenance / process aspect (how the conceptual and physical creation came into recognisable existence; and how we know what perceptible “thing” to look for);
2. The abstraction aspect (how the intellectual content, as distinct from its form, medium etc., is recognised, identified and described).

These two aspects will be elaborated further in the sections below and compared across the cultural heritage and cultural industries sectors.

5.3.2 Resource or copyright content focus?

Rust (2005a) exhaustively described the main options for responding to the second of these core problems (the abstraction or “work” problem):

“there are three different common uses of the term “work” in metadata systems which are often confused and are in reality quite different things:

- (a) a “legal” work, of which there are different definitions according to different jurisdictions;
- (b) the “preconceived” work, as imagined in the mind of the creator – “what the author intends” in a creation;
- (c) the “abstracted” work, according to any party who has reason to identify a conceptual “work” underlying a particular **manifestation**.”

The emphasis on **manifestation** is added so that the different ways of seeing a resource in the cultural heritage and commercial content sectors may be seen at the clearest point. As noted in section 4.3. Conceptual Frameworks, cultural heritage tends to view unique items (or semi-unique because selected as an exemplar!); these are the “manifestation singletons” identified in FRBRoo. The classes of identical products essential to the commercial content view is the FRBRoo “manifestation product type”⁴⁹. Although this is a stark generalisation, this distinction explains the difference in focus between sectors (although all three uses are understood and acknowledged in commercial and heritage settings):

Sector	Way of seeing “Resource”	Type of “Work”
Cultural heritage	“this particular manifestation” of a cultural object e.g. the Mona Lisa hanging in the Louvre (c.f. “manifestation singleton” below)	The process of creation, cultural influences and impact, and provenance of this particular object e.g. historical, cultural, critical etc. views of da Vinci’s preconceived depiction of the Mona Lisa

⁴⁸ This relationship of “who says” may be represented by introducing “reification” for example:

<http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#Reif>

⁴⁹ See FRBRoo: http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_2009_june_.pdf

Sector	Way of seeing “Resource”	Type of “Work”
Commercial content industries	“the class of all identical products manifesting” this object e.g. all the A3 posters of the Mona Lisa in the same format, medium, price etc. (c.f. “manifestation product type” below)	The abstracted work identified in all identical product type manifestations e.g. the posters by any party involved; for example the Louvre museum, photographer, publisher of the poster (probably not da Vinci’s estate since the 1500s)

5.3.3 How metadata is used in public cultural heritage sectors

The five objectives of a library catalogue system, which could perhaps be generalized to objectives for any heritage sector documentation system⁵⁰, were enumerated for the 21st Century context by Svenonius (2000) as:

- finding;
- collocating;
- choice;
- acquisition;
- and navigation (added to the original four “user tasks” noted in FRBR⁵¹).

All of these objectives apply to some extent in one or more levels of the supply chain for most or all cultural industry sectors. However, here we need to add many more objectives for information systems, and thus metadata, specific to the commercial sector:

- identifying (disambiguating and authoritatively checking);
- version tracking;
- notifying availability to market;
- pricing and trading;
- recording and protecting legal rights;
- syndicating;
- aggregating;
- distributing;
- marketing and promotion / differentiating from competition’s offerings;
- sales reporting;
- calculating and distributing royalties.

5.3.4 How metadata is used in the commercial cultural industries

Commercial metadata, in sharp contrast to (most) public sector metadata, are not static repositories of data but rather *data flows*.

An outline of the supply chain partner needs met by commercial sector metadata, as against the uses of public sector heritage organisations will help to characterise the object of this research. It should be noted in the table below that the use cases in the commercial sector column tend to feed into those for the next step in the supply chain, sometimes interrelating with previous steps, whereas those for the public sector do not; they tend towards a more static repository of data.

⁵⁰ See Appendix 1 – Glossary of Terms for a definition of cultural heritage “documentation”

⁵¹ FRBR, 2009. p. 79.

Type of metadata [examples]	Public cultural heritage	Commercial cultural industries
Authoring and publishing workflow [contributors, change logging, version numbers, keywords, formatting descriptions, audience codes, project management information]	Possible uses for Open Education Objects	Preparation, formatting and publishing original cultural content
Descriptive [subject/topic descriptors, genre, context, contributors, relations, language]	Identifying and discovering content for ordering, acquisition, ingestion and discovery by users	Identifying and discovering content (owned or licensed) at point of creation (or acquisition)
Administrative [creation of record, creation of digital object, technical requirements, internal identifiers]	Essential to management of collection using (more or less) standardised information systems	Normally created and maintained in-house within highly customised information systems
Holdings [numbers of items, locations, availability status]	Number of items held for use, location and access arrangements	Analogously, quantities supplied or available to market
Technical [software and data formats, physical carrier format]	Carrier format(s) Access practicalities for users Preservation	Carrier format Cost of bringing product to market versus possibility of reaching users of a particular technology platform
Preservation [all of the above plus aggregation/provenance preservation schedule/workflow information]	Ensuring future discovery and access by users Creating cultural / scholarly record	Internal archiving and/or “Digital Asset Management” processes to enable long-term exploitation and care of IP (intellectual property)
Structural [component work identifiers, semantic tags, reading flow indicators, display formatting]	Enabling discovery of semantically relevant parts of a resource Supporting archival needs	Maximising exploitation of IP by offering relevant parts to market Identifying interested parties with rights in parts of larger works

Type of metadata [examples]	Public cultural heritage	Commercial cultural industries
Rights / licensing [provenance/authority of rights claims, rights descriptions, parties identifiers, contractual terms, jurisdictions]	Ensuring legal compliance (copyright and licensing) Financial probity and transparency	Ensuring legal compliance (copyright, licensing and contractual rights) Financial probity and transparency (royalty computation) Maximising return on investment for IP
Pricing [price, currency, market/territory, date]	Not pricing as such as service cost is more relevant Acquisition costs recorded for accounting and strategic decision making	Ensuring competitive offerings Return on investment Availability status and order point
Marketing [collateral/previews, representative extracts]	Possible (indirect) analogy with marketing services, information literacy initiatives	Differentiating product from similar offerings within market Maximising sales
Sales [usage/download statistics, dates, IP ranges]	Strategic decision making (value for money) Analogous data would be usage statistics, possibly costs incurred for services like inter-library loans, paid access to a database service, use of a physical object, copies of archive materials <i>etc.</i>	Strategic decision making (return on investment) Accounting, product management, royalty computation and payment

Table based on ODL, 2005; mEDRA, 2002

5.3.5 Metadata best practice

“Best practice” includes the use of standards, adherence to expert recommendations and research. This definition holds even when a standard is not yet widely used (the EIDR, for example – see section 6.5.2) but still represents the industry’s recognised highest expertise. A key difference in commercial sector metadata that is general enough to merit separate discussion is the issue of metadata best practice above and beyond the implementation of a given standard metadata schema or identifier; by this is meant how, when and by whom metadata are registered for or added to a resource, first communicated to a supply chain or updated. These data flows need to be managed according to clear guidelines.

Depending on the industry sub-sector's needs, and those of the particular supply chain partners, a variety of best practices will be recommended (and followed to varying degrees). However, it is possible to make some general statements about the nature of these practices using the clearest examples.

A general statement covering all of the following is “that metadata should be created once, used many times for different purposes”⁵². Very much unlike in the cultural heritage sector, the creation of metadata should happen at the very beginning of the creative process, even if possible before the stage of perceptible manifestation of the creative concepts.

The Metadata Working Group “Guidelines for handling metadata: version 2.0” (MWG, 2010) lays out a code of practice or “actor model” for metadata creators, handlers (“changers”) and receivers (“consumers”) at each point of the chain:

Creator	• creates the <i>first</i> instances of metadata in a new file • <i>must</i> have full knowledge of all metadata it is creating • <i>must</i> always create standard compliant metadata in at least one form
Changer	• may <i>only</i> delete metadata with specific intent. • <i>should</i> obey rules for <i>consumer</i> applications when reading metadata. • <i>should</i> keep all forms of metadata it modifies in sync with one another
Consumer	• <i>only</i> reads metadata from the file • <i>may</i> use metadata for display purposes, searching, content organization, <i>etc.</i> • <i>never</i> modifies the metadata in the file itself. • <i>must</i> reconcile between different forms of metadata in the file • <i>must</i> use metadata according to the semantics defined for each field

Table adapted from MWG, 2010, pp. 15 – 17.

The MWG model of metadata use was developed for human actors and (semi-)automated processes built into software applications, specifically in the still images sector. However, it is general enough to use for any metadata supply chain (at least for most varieties of metadata) and since it is maximally prescriptive it acts as a useful benchmark for other codes of practice.

Schema validation

Also important for the exchange of messages between automated systems, and the minimisation of checking large amounts of data by hand, is the technical validation of messages against the relevant schemas. This refers to the checking of the presence, correct syntax and allowed content of the metadata elements in an instance of a particular kind of data, against a schema to ensure that only conformant records are accepted and errors of syntax and semantics can be flagged.

In practice this will usually mean validating an XML document against its DTD⁵³ or schema. A DTD specifies only the structure of the data elements within the document; the schema goes further by allowing also automatic checks of allowed values within each data element. The limitation of these

⁵² See Indexes, 2000, p. 8.

⁵³ See the W3C description of XML validation: <http://www.w3.org/TR/xmlschema-1/#concepts-schemaConstraints>

automated approaches is that they do not address the meaning of the whole document, created by combinations of elements; e.g. they will not detect that a hardback product form does not apply to an ebook, even if both these facts are, independently, correctly described.

5.3.6 Standards: Origin, adoption and maintenance

Best practice can be developed and recognised within one company, based on academic research, promulgated by experts and consultants, and very often, analysed and formally expressed in some kind of standard.

Standard identifiers and metadata frameworks arise in the commercial sector broadly speaking through two routes: *de jure* standardisation or industry coalition. Although *de jure* standards can be introduced by bodies at various levels, for example, at the European level by the European Committee for Standardisation (CEN), at the national level by a large number of bodies (e.g. BSI in the UK; DIN in Germany). Although for cultural heritage, bodies like NISO in the US have been influential (e.g. with the Z39.50 standard), the most relevant body in this context is ISO (International Organisation for Standardisation).

The ISO and trade standard routes are not mutually exclusive; an industry standard may later be proposed and accepted as an ISO standard. ISO's own definition of a standard states in this regard only that it should be "established by consensus and approved by a recognized [*sic*] body"⁵⁴, so there is a good deal of common ground between the two approaches.

Internal industry standards

An industry standard is developed through deliberate collaboration between several organisations, often (but not always) at different points in one industry's supply chain. One or more commercial companies become aware that rationalisations could be made in their internal operations, relations with clients and suppliers, and reporting mechanisms, resulting in either more efficient business processes, cost savings, economies of scale, or all of the above; that is, for a demonstrated economic benefit.

Delegates from the major players are appointed to consult on meeting the need for standardization and proposals discussed before adoption by committee. The standards are then implemented by each partner and maintained by a cross-sector group, often supported by subscriptions from members (who may increase beyond the original collaboration as more companies see the value of the shared standard).

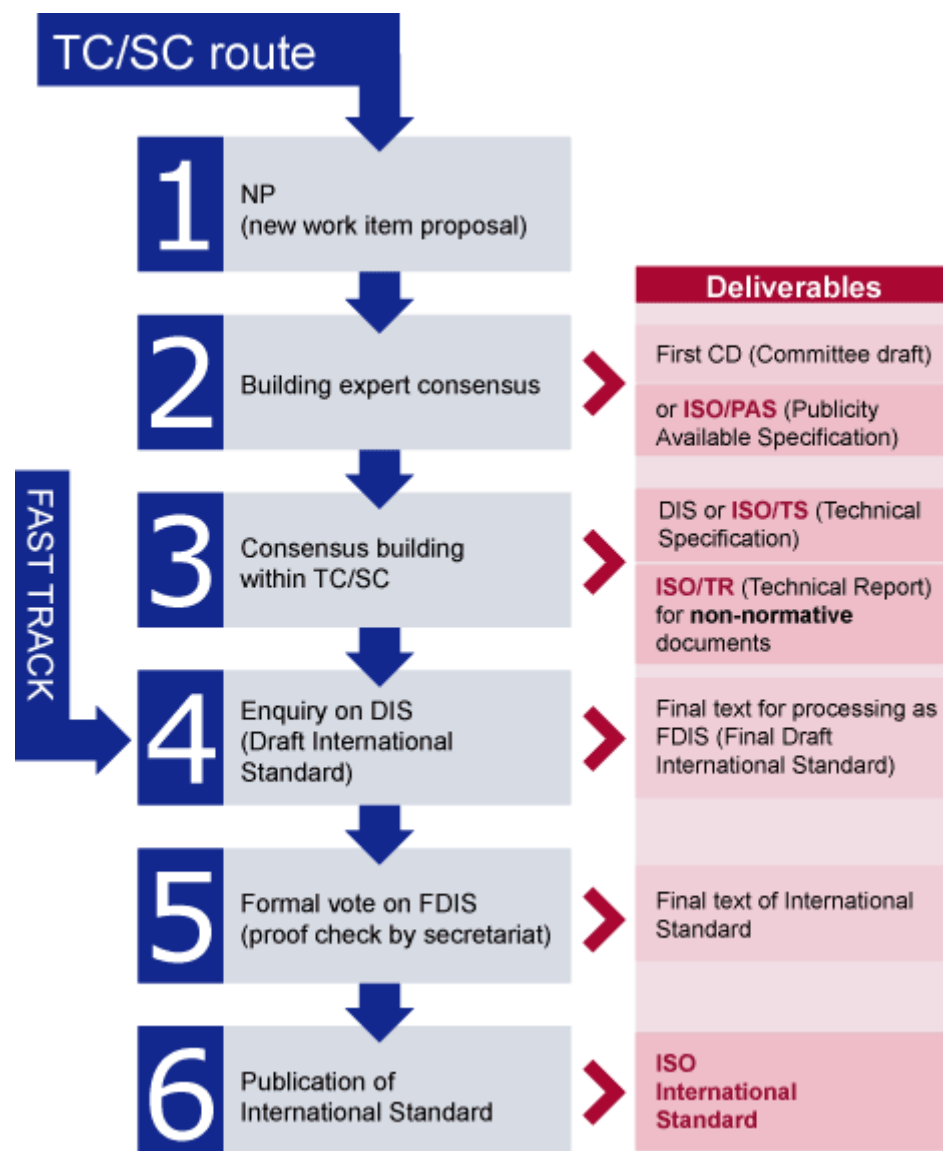
ISO standards

The International Organisation for Standardisation route is clearly the more formal process⁵⁵, and is dependent on successful voting at various stages by member countries.

⁵⁴ See ISO, 2011

⁵⁵ For more details see:

http://www.iso.org/iso/standards_development/processes_and_procedures/how_are_standards_developed.htm



Schematic representation of ISO deliverables⁵⁶

This encompasses three phases: work item proposal (including acceptance of the need for a standard, agreement of scope); preparation of a draft and negotiations of specific technical details; approval of a final draft.

After publication, a review of the standards usually takes place at least every five years, and, at least in the case of the identifier and metadata standards here described, this might coincide with official approval of the agency appointed to administer the standard practically.

5.3.7 Commercial sector identifiers

An identifier in general terms is simply a label used to recognise someone or something in a given context. In daily life, these can include titles, qualified nouns and people's names (see Rust, 1998: "there is ultimately no high-level logical distinction between a title and an identifier"). However, in any complex situation, and especially where legal and commercial commitments are involved, identifiers need to be carefully controlled to make them unique and unambiguous.

⁵⁶ From the ISO website:

http://www.iso.org/iso/standards_development/processes_and_procedures/deliverables/deliverables_schema-2.htm

Essential requirements of identifiers

An identifier itself should be persistent, unique and system independent. These are minimum characteristics, recognized by the cultural heritage sector and the commercial sector alike. As has been noted (McKenna & Wyns, 2010) these objectives are not always achieved in practice.

The Indecs (2000) framework notes that “the most important properties of an identifier are

1. uniqueness within a given domain;
2. stability (identifiers should never be changed or transferred);
3. security, whether through protection by watermarking or encryption, and/or by internal consistency through the use of check digit algorithms; and
4. the public availability of some basic descriptive metadata for the entity identified, without which the identifier has only limited use” (Bide and Rust, 2000).

System independence is a feature of the standardisation process itself. Uniqueness and persistence are guaranteed by the registration of products by the identifier’s managing agency. A reasonable level of “intelligence” is useful to make managing the identifier practical but in theory is optional, as are extra services, beyond mere identification, built around a registry.

Finally, it should be noted that the requirement for genuine uniqueness entails a minimum set of description data (possibly including other identifiers, for example for components or related parties) that may itself form part or all of another standard. This metadata needs to contain all indispensable data for each supply chain partner to make a commercial and economic decision.

Requirement	Mechanism	Explanation
Persistence	Registration: Stability of registry (and its maintaining organisation) Error detection mechanism	The identifier is stable, theoretically permanent, and reliable across all supply chain stages
Uniqueness	Registration: Reference descriptive metadata No deletes No reuse	An identifier is only reliable if it consistently points to only one referent. Descriptive metadata allows disambiguation and removal of duplicate registrations, and best practice ensures avoidance of multiple referents
System independence / interoperability	Published standard syntax and format	It must be possible to implement in diverse organisations with different IT infrastructures and working practices

Requirement	Mechanism	Explanation
Meaninglessness – “unintelligent” / “low affordance”	Use of encoding algorithms and tables within registration service	The lowest affordance IDs are “dumb numbers” – in fact a random number would be best, as a serial number contains information about the relative temporal order of registrations – but some level of reference to a familiar concept helps reassure users that registration is coherent Encoding a large amount of metadata may risk exposing commercially sensitive information or more significantly, risk misapplication of implied intelligence
Actionability	Paid, trusted third party managers of service infrastructure and maintenance	Added services around identifiers are optional but increasingly popular in digital content businesses

Governance of a standard identifier

Although governance (high-level, policy decision making) and management (e.g. registration and search provision) of standards vary across sectors, the basic features are the same. There is normally a hierarchical organisation including:

1. Registration Authority – the governing body, implementing top-level practical policies and managing the lower level of agencies;
2. Registration Agencies – administering territories, markets or other groupings of registrants.

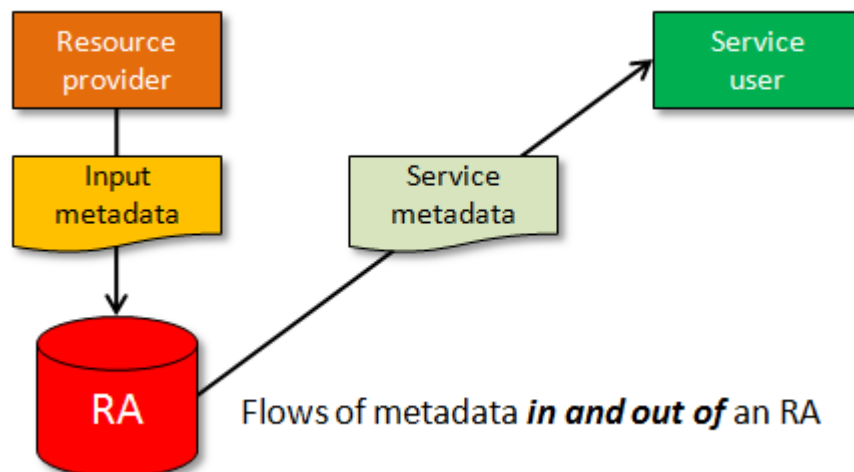
In many cases (but not necessarily) the Registration Authority is the proposer of the current version of the standard in question. In most cases they will also be responsible to a board of directors drawn from the industry served.

The precise division of tasks and responsibility between Registration Authority, Registration Agency and Registrants, and the levels of read or write access they have to each other's data in particular situations varies significantly across industry sectors. Hence it is difficult to generalise about the governance aspect.

Infrastructure – registration and reference minimum metadata

An extremely important aspect of the governance of a standard identifier is the registration of *minimum reference descriptive metadata*. This is the set of mandatory data elements required by the registration agency to allow registration of each entity, whether at a work, expression or manifestation level. It has two functions:

1. To enable unambiguous registration and look-up of entities directly from the registration agency (creating and retrieving the identifier of a given entity);
2. To enable the industry supply chain to function reliably along its entire length, on all branches (re-using the identifier).



http://www.doi.org/handbook_2000/metadata.html

The diagram above, from the International DOI Foundation (2011), shows this principle at its most basic level, and shows the distinction between the minimum reference data (here “input metadata”) used to register a resource, and the “service metadata” which may be more extensive depending what services are associated with registration, provided by the registration agency.

The minimum case above is also the most basic possible example of a *service*⁵⁷ associated with an identifier system. Services accompany many identifiers in the commercial sector and as they depend on technological infrastructure and human intervention, naturally must be paid for at some point, as a cost to the party registering the identified resource, or (and in some cases also) by interested parties wishing to gain access to the related information. This is one of the key reasons behind the intrinsic commercial nature of this data (see section 3.4 on the Europeana DEA and section 7 on legal-commercial implications).

*Choreography of messages*⁵⁸

The example of registration and look-up of an identifier for a given entity is the simplest possible example of a message choreography. This is a description of possible messages, normally based on notifying business partners of one or more identifiers for products or other entities (e.g. market offerings, interested parties *etc.*) but also including responses, corrections and additions. The existence of pre-defined messages underlines the dynamic nature of commercial metadata as data flow. Their existence is also an example of best practice in terms of necessary business communications (choreographies apply especially in situations involving rights and legal claims) and so also underlines the essential legal-commercial nature of this data.

5.4 Conceptual models for data architecture and interoperability

Conceptual models have two uses in the context of metadata and identifier best practice:

1. Data architecture:
Creating identifier systems that refer to entities at the correct level of abstraction, and metadata that take into account the relevant attributes and relations;

⁵⁷ See Appendix 1 – Glossary of Terms for a detailed explanation

⁵⁸ See the W3C definition of a generalised choreography for any type of data:
<http://www.w3.org/TR/ws-cdl-10/#Choreography>

2. Interoperability:

Mapping genuinely semantically equivalent entities; neutrality across hardware and software; creation of effective intermediate mapping, storage and transmission schemes.

All conceptual models for “information objects” or “cultural objects” have in common a basic distinction between the conceptual (intellectual) and perceptual (sensory) aspects of any type of object: “there has been a general consensus that stuff can be divided into two kinds of thing. This dichotomy has been described in many ways – ideas/things, abstract/physical, concept/percept. The basic dualist approach is largely uncontested...” (Rust, 2005a).

*The Indecs, CIDOC and FRBR(oo) models*⁵⁹

As this work package aims at building common understanding between the commercial and public sectors, it uses technical terms taken from both, and shows their interrelations. This is desirable in order to allow technical workers in both fields to appreciate the possibilities for collaboration. In order to see how the terms relate to each other, it is necessary to understand something of the underlying conceptual frameworks that inform both sectors’ professional terminology.

5.4.1 CIDOC Conceptual Reference Model

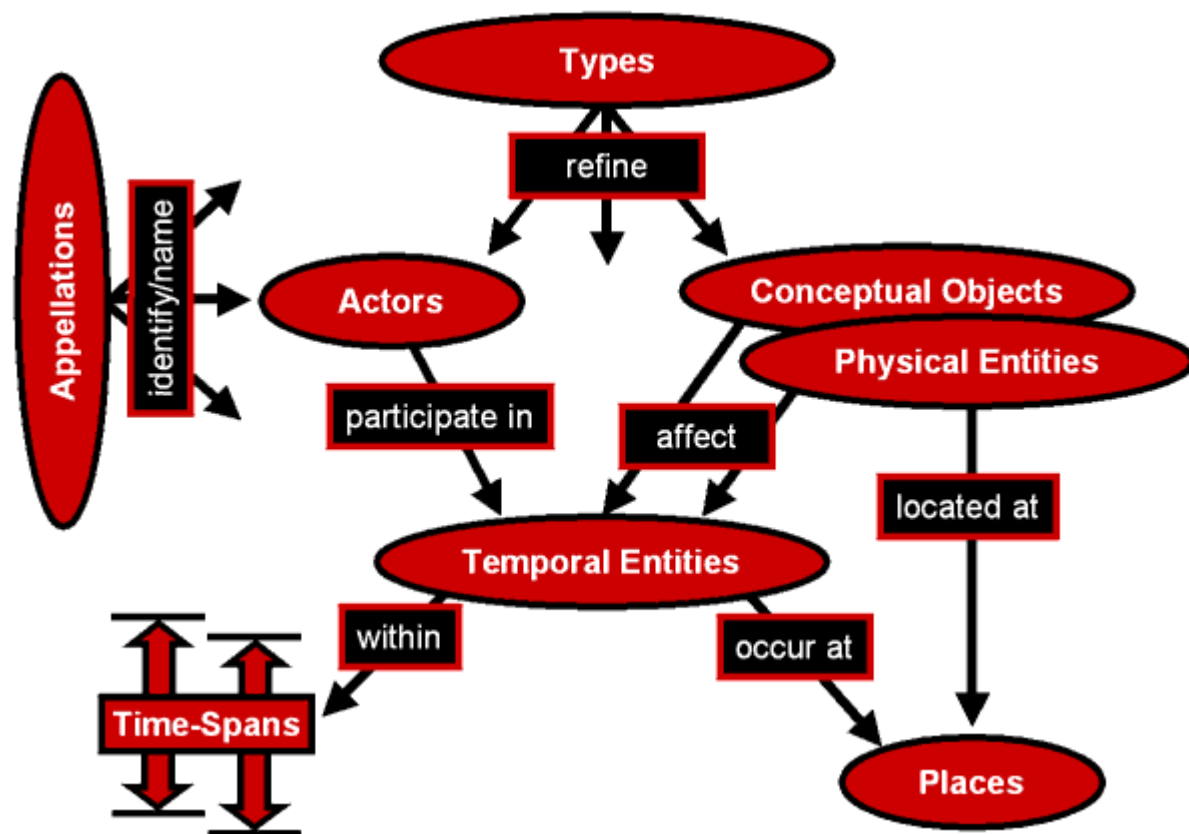
CIDOC-CRM is the ontology developed from international standards in the museums domain for describing the entities underlying the varied information types produced by museums⁶⁰. Although it is a reference model, and not intended as a blueprint for metadata schemas or information systems, its relevance to Linked Heritage is that it aims to ensure interoperability between disparate museum information systems, and thus informed the design of LIDO, the metadata harvesting format developed by ATHENA and used to populate Europeana from the heritage sector (see Doerr, M. *et al.*, 2010).

Of the three models compared here, CIDOC’s Conceptual Reference Model has the most component parts and arguably the most formally rigorous overview of cultural and creative activity. As the diagram below shows in extremely simplified form, the central organizing entity is the event, and this allows processes and complex creative works (material and immaterial) to be modelled in detail.

Rust (2005a) notes that this event and context-based approach is held entirely in common with Indecs. Although CIDOC-CRM’s focus is naturally on description of artefacts (“items” or at least “manifestation singletons”), “whereas indecs is most concerned with the manifestation/abstraction relationship as it applies to classes” (Rust, G., unpublished comment, 2011), it is this underlying commonality that will enable mappings.

⁵⁹ Special thanks are due to Godfrey Rust in significantly revising sections 5.4.1. to 5.4.3.

⁶⁰ See Crofts, *et al.*, 2010.



CIDOC overview diagram quoted from Gill, T., 2004.

5.4.2 FRBR(oo)

FRBR is a conceptual model of the objects of library catalogues and the types of statements made about them by cataloguers in order to allow fulfilment of the “user tasks”: finding, collocating, choice and acquisition (FRBR, 2009). Since it was originally developed FRBR has been criticised, especially because its “flat” approach to creative works leaves the boundaries between definitions of entities unclear (as Carlyle, 2006 explains) and did not address the questions of the publishing process and complex works and expressions. These were addressed and “solved”⁶¹ by harmonising FRBR’s terms with the ontological, event-based approach of CIDOC-CRM, resulting in the FRBRoo “object-oriented” ontology⁶².

FRBR is an entity-relation model. The top-level entities in FRBR are centred around the various ways in which a creative “work” is made perceivable (as expressions, manifestations and items), and the remaining components of the model (the subjects of works and the agents who create and express it) are related in various ways to these.

The important aspect to note here is that FRBR’s model is centred on describing static cultural entities in one dimension (degree of abstraction). Because it lacks a description of the processes that gave rise to these entities (the “event model” which is common to CIDOC and indecs), it is unable to readily distinguish between types of manifestation and expression important to the commercial sector (and even to some extent to the cultural heritage sector).

⁶¹ See the FRBRoo presentation “Modelling Intellectual Processes” at http://www.cidoc-crm.org/docs/frbr_oo/FRBR_tutorial_gothenburg.ppt

⁶² See FRBRoo section at the CIDOC website: http://www.cidoc-crm.org/frbr_inro.html

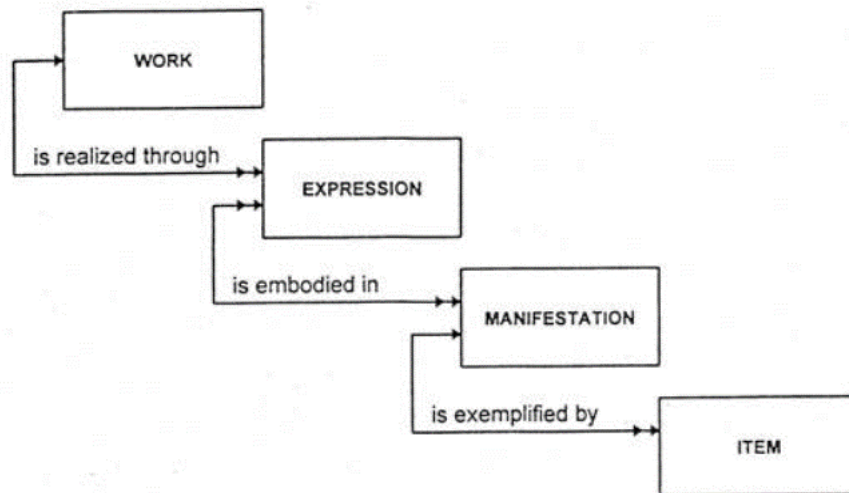


Diagram of the FRBR top-level entities (see IFLA, 2009).

Although the real world entities derived from creative works in the FRBR and Indecs model are the same, the ways in which they are categorized have significant (though mappable) differences:

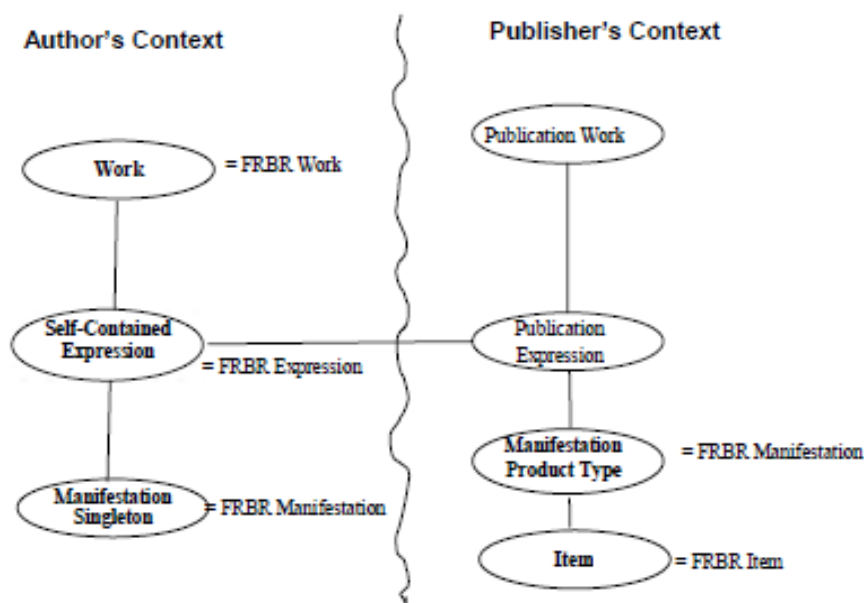
- The FRBR “work” and FRBR “expression” are different degrees of an indecs “abstraction” (and in practice the indecs abstraction is often called a “work” in specific schemes such as DDEX or ONIX);
- The term “expression” in Indecs means something quite different from that in FRBR – it is a spatio-temporal expression or performance, in contrast to a “fixed” manifestation. FRBR recognizes these as a type of FRBR expression;
- “Manifestation” has the same meaning in both;
- The FRBR “Item” was not made explicit in Indecs but is compatible with it as an instance of a manifestation.

The harmonisation of FRBR with CIDOC which resulted in FRBRoo, an ontology which shares much more of the concerns of Indecs and provides a conceptual bridge between the galleries, museums, and archives, and the cultural industries. Because FRBRoo was concerned with solving particular problems in these areas, it has much more value for Linked Heritage than more abstract interoperability projects like ABC⁶³, even though they might have valid insights for mapping purposes.

The main insights of FRBRoo directly address the key difference between these two “worlds”:

- Commercial sector “manifestations” are a class of items, not a unique item. Most available commercial metadata describes this “product” class. FRBRoo calls this the “manifestation product type”;
- Cultural sector metadata describe “manifestations” which are generally unique items, and these FRBRoo calls “manifestation singletons”.

⁶³ See the ABC Harmony Data Model Version 2 at: <http://metadata.net/harmony/ABCV2.htm>



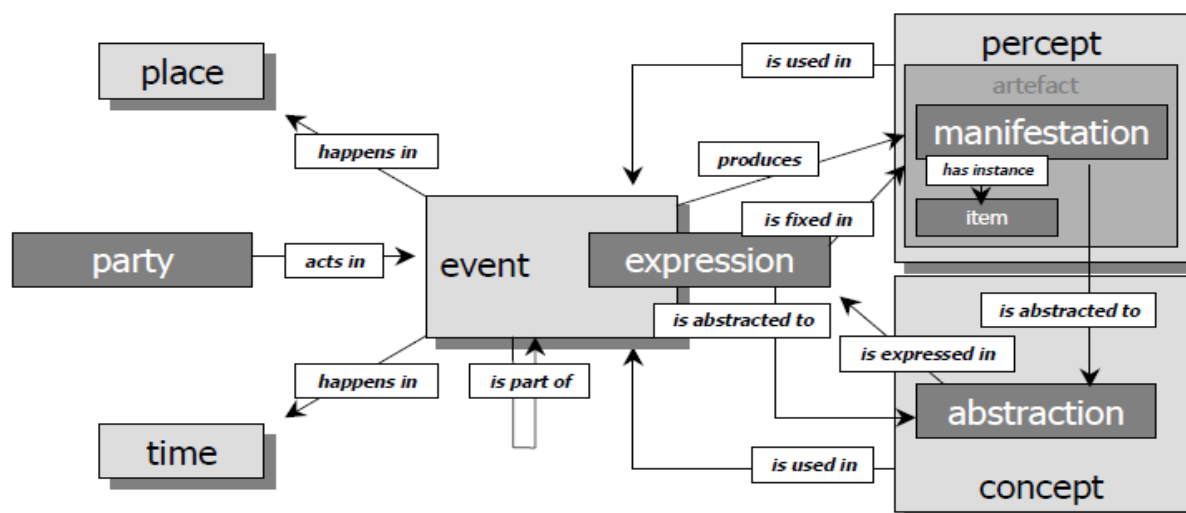
FRBR(oo) analysis of authoring and publication views of creating a book⁶⁴

The underlying models of creation and publication (or dissemination) of these two types of manifestation contain similar entities, actors and relations to the Indecs “model of making” described below.

5.4.3 Indecs Metadata Framework

In the Indecs model there is a clearer focus on the events and relationships involved in the creation, expression and manifestation of works and the relationships created between actors, due to the commercial view taken, the rights tracking needs of commercial actors, and the need to deal with the processes involved in mass replication and dissemination. The core processes of the Indecs framework are described in the three events “people make stuff”, “people use stuff”, “people do deals about stuff”. Of these, the first, known as the indecs “model of making” and illustrated in the diagram below (Bide and Rust, 2000), has the most direct overlap with FRBR:

⁶⁴ From http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0_2009_june_.pdf



The events-based approach is strikingly similar to CIDOC-CRM, although all three of the models referenced here were developed quite independently. This reflects their common interest in the way resources come into being and are used. However, the similarities are only partial; CIDOC-CRM deals mainly with specific, more-or-less unique *items* whereas Indecs is naturally more concerned with *manifestations* or mass-produced *products*.

Any mapping between the two has to take into account this shift by one degree up or down the scale of abstraction, as well as the richness of provenance detail in CIDOC-CRM as against the proliferation of use and re-use terms and conditions in Indecs. However, there are no overwhelming obstacles in reconciling these models, or schemas which have made use of them, and the three models can be seen as usefully complementary in terms of broadening the scope of available metadata rather than in any way in conflict.

In the rest of the report, terminology from the Model of Making above will be used throughout. Refer to Appendix 1 – Glossary of Terms for the detailed definitions and equivalents of terms in other conceptual models.

5.5 Identifiers and metadata frameworks

The data available in the commercial sector may be categorised under two main headings and several sub-categories as in the table below, due to the differences in their function and behaviour.

Categories of data:	Standard identifiers	Metadata frameworks
Sub-categories:	<u>Non-actionable</u>	<u>Standardised</u>
	<u>Actionable</u>	• <i>de facto</i> • <i>de jure</i>
	Internal (proprietary)	Non-standardised (proprietary, technically-imposed or <i>ad-hoc</i>)

Categories of data:	Standard identifiers	Metadata frameworks
Uses:	<u>Supply chain</u>	<u>Supply chain</u>
	<u>Trading</u>	Reference descriptive metadata for identifiers
	Royalties payment	Rights information harmonization
	Sales tools	<u>Marketing</u>
		Sales reporting

The sub-types and uses underlined in the table above are those most obviously important for Europeana and thus are emphasised in this report. The distinction between proprietary and non-proprietary identifiers (those used purely within an enterprise, versus use along the whole supply chain) is particularly important because “best practice” in this report refers to usefulness across all partners in a supply chain.

Even if a great amount of commercial data is available in proprietary formats (e.g. Europeana contributions from The Orchard⁶⁵ and WOMEX⁶⁶), proprietary identifiers will not necessarily be resolvable outside their original organisation, and proprietary metadata may be semantically uninteroperable. However, for the future deliverable on technical framework may need to take them into account as this may be all that is available from some metadata services.

What type of data is recorded, transmitted and stored?

It must be emphasised again that the use cases for commercial sector product identifiers and metadata have some aspects that are fundamentally different from those for non-commercial heritage sector standards.

Heritage organisations commonly accept that persistent identifiers for digital and physical objects, and interoperable metadata standards⁶⁷, are essential to their public in a complex networked “information ecology”, and thus have many of the same goals in creating and publishing metadata as cultural industries.

However, the commercial industries must reckon with satisfying the requirements of supply chain partners including the creators of cultural works themselves, and this involves different moral, legal and technical constraints, such as the moral rights of creators to recognition and attribution of their efforts, their fundamental legal rights (e.g. copyright) and those created by contracts at each stage of supply to the public, and the economics and technological aspects of building and maintaining supply chain infrastructure.

As calls for mass digitisation of in-copyright, out-of-commerce works increase⁶⁸, the need to form public-private-partnerships across these two domains, and the concomitant need to deal with the

⁶⁵ See Europeana search:
<http://europeana.eu/portal/search.html?query=the+orchard&tab=sound&qf=TYPE:SOUND>

⁶⁶ See Europeana search:
<http://europeana.eu/portal/search.html?query=womex&tab=sound&qf=TYPE:SOUND>

⁶⁷ See ATHENA deliverables D3.2 and D3.4: <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>

⁶⁸ See recent expressions such as <http://www.ifrro.org/content/stakeholders-sign-groundbreaking-mou-out-commerce-works> and <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/619&format=HTML&aged=0&language=EN&guiLanguage=en>

issues familiar to the cultural industries, will become more urgent for the cultural heritage sector. Hence the concepts in this report will become more valuable to the cultural sector with time.

5.5.1 Overview of identifiers used in the cultural industries

The following table (adapted from International ISMN Agency, 2008) shows the major commercial sector identifiers (all standardised, except the PLUS scheme, which is still in a pre-standardisation state). Note that the ISSN used for serials is included here for completeness but is out of scope for this report. Each identifier has been aligned with the kind of entity it identifies, according to the commercial Indecs classification and the heritage sector CIDOC classification. It should be noted that DOIs and URIs can identify entities at *any* level of abstraction in either sector, and that some schemes, being new or only proposals, are not actually used as yet.

The identifiers underlined will be discussed in depth in this report.

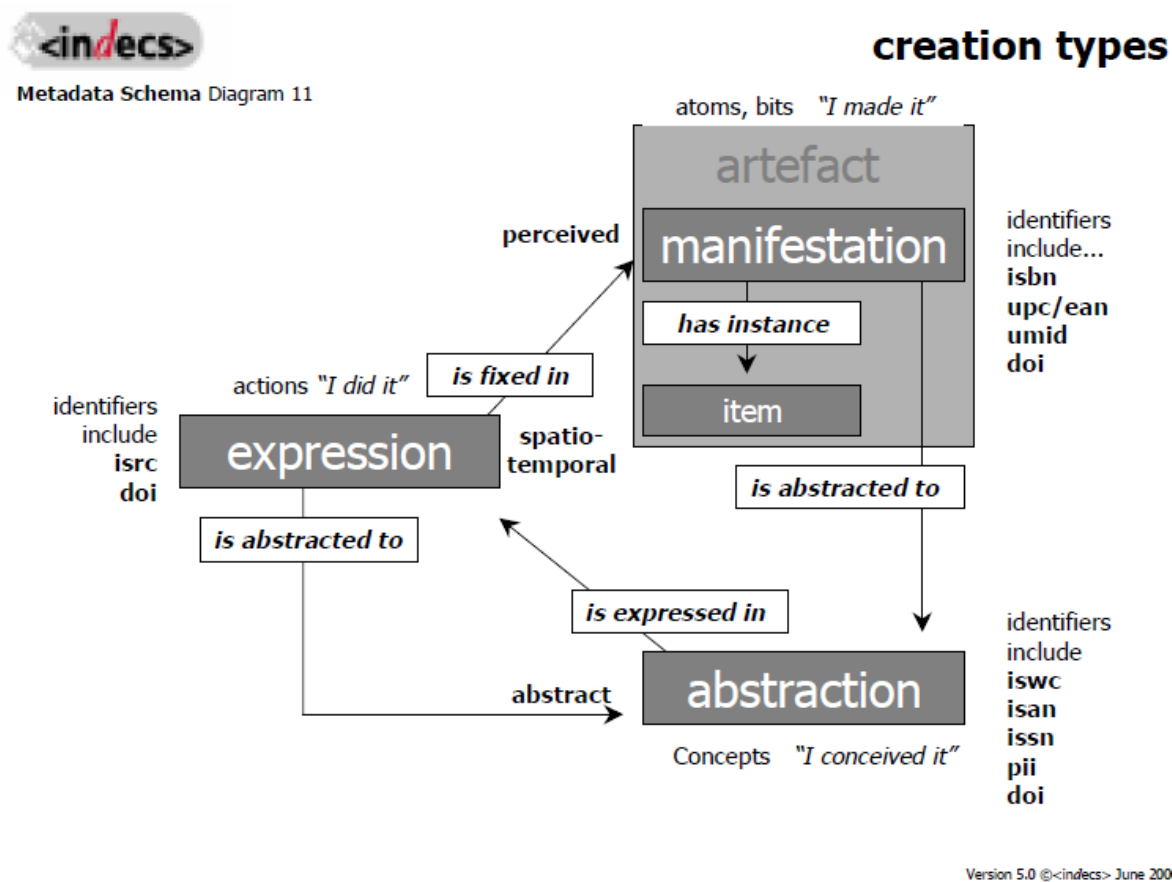
Identifiers	“Real” entities identified	Indecs type identified	CIDOC type identified	Who can use this level of identifier	Level of abstraction
DOI, URI – can identify any kind of entity Commercial, domain-specific implementations include: EIDR (AV), ISBN-A (books), CrossRef (academic journal articles)	<u>ISNI</u>	Names – “public identities”	Name (attribute)	E82 Actor Appellation	Creators / authors / performers / fictional characters / brand names
	[MWLI] [PLUS ID]	Contracts [and all other legal entities within eventual PLUS framework]	Agreement (event)	E5 Event	Music publishers / Picture libraries
	<u>ISWC</u> <u>ISTC</u> <u>ISAN</u>	Musical compositions / Texts / AV productions	Abstraction (entity)	E73 Information object E33 Linguistic object	Rights societies / publishers / AV producers, distributors, retailers
	<u>ISRC</u>	“Recordings” [independent of carrier form]	Abstraction (entity)	E90 Conceptual object	Producers
	<u>GRid</u>	“Releases”	Abstraction (entity)	E90 Conceptual object	Music Distributors
	<u>ISBN</u> <u>ISMN</u> ISSN (GTIN-13)	Books / Music scores / Journals [as generic products – i.e. the set of all identical products]	Manifestation (entity)	E31 Document [E36 Visual item]	Wholesalers / retailers
	None	Items [instances of the products in the row above]	Item	E84 Information carrier	Not relevant to mass production industries (although maybe to customer support?)

Abstract Concepts

Concrete product

Items

The Indecs framework (Bide and Rust, 2000) also contains a helpful map of the creative process (essentially the “Model of Making” described above) showing the types of entities identified by some of the standards described here:



As stated above, these terms will be used to denote the referents of the identifier standards here described.

5.5.2 Overview of metadata schemas used in the cultural industries

The table below shows the metadata schemas described in this report, with the identifiers used to specify the referent product or abstraction described by them.

Industry sector	Book publishing	Recorded music	Film and TV	Images
Metadata schema	ONIX for Books	DDEX	EIDR	IPTC PLUS XMP
Links to Identifiers	ISBN / ISMN ISBN-A ISTC	ISRC GRid	ISAN EIDR ID	PLUS ID

There are very few published overviews covering all of these sectors and standards. The current W3C “Ontology for Media Resources” appears to share some of its scope⁶⁹. One project which previously reviewed the main schemas in all sectors was the highly technical Vocabulary Mapping Framework (VMF)⁷⁰. A recent review article (Bide, 2011a) has also taken an overview, from a holistic, supply chain, legal-commercial and technologically integrative perspective. However, this report appears to be the only published attempt to present detailed descriptions, analyses and syntheses of all of these standards, adding to the originality of this research.

⁶⁹ See <http://www.w3.org/TR/mediaont-10/>

⁷⁰ See <http://cdlr.strath.ac.uk/VMF/> as well as <http://www.doi.org/VMF/documents.html> and <http://www.doi.org/VMF/archive.html>

6 Industry Sector Findings

The results from this investigation have been touched on throughout [the first 3 sections] of this report, in order to illustrate the concepts involved.

Now each cultural industry sub-sector is addressed in turn and all the standard identifiers and metadata schemas, along with their interrelations [and other interesting frameworks] described in detail.

6.1 Presentation of results

In compiling this report the aim was to describe the frameworks used in each commercial sub-sector in the same way, in order to bring out the similarities as well as differences in standards and approaches. Where this has not been realistic, the uniqueness of the situation has been highlighted and possible explanations offered. The top-level aspects described are as follows:

Presentation of identifiers	Presentation of metadata schemas
Introduction	Introduction
Scope	Scope
Governance	Governance
Construction	Construction
Reference descriptive metadata	Other versions and related data models

Metadata schemas are described using a standard set of categories developed as a first attempt to bring together the common practice of the commercial sector metadata standards landscape as a whole. The report by mEDRA (2002b) covered many of the same schemas and standards, as did Altomare (2007) and EDItEUR and BNF (2010); the formats used by these is almost always followed here as their scopes and terms of reference are compatible with those of Linked heritage.

Identifiers' construction is described using a standard notation: first the acronym of the identifier's name, then its structure, separated into elements, where they exist, by hyphens. This is done for clarity of explanation, even where the standard deprecates the use of hyphens in presenting the identifier.

The letter “**d**” will stand for a *decimal* digit (*i.e.* numbers 0–9), “**a**” for *alphabetic* only, “**c**” for any alpha-numeric *characters*, and an “**h**” for a *hexadecimal* digit (*i.e.* a base-16 digit expressed using the numbers 0–9 then letters a–f). This is based on, but not limited to, practices common to the various identifiers across all sub-sectors. Where identifiers are made up of elements that encode some meaning⁷¹ *at the time of registration* (but which may not apply at future times) these are tabulated for clarity, but should not be confused with metadata elements intended for later interpretation:

Element	Length	Possible values	Meaning
---------	--------	-----------------	---------

⁷¹ See <http://www.bic.org.uk/files/pdfs/uniqueid.pdf> for more on intelligence of identifiers

Identifiers' Reference descriptive metadata are included in the identifier's description as they are normally required to register an instance of the identifier:

Data Elements	Cardinality	Allowed values	Comments
---------------	-------------	----------------	----------

The term “cardinality” above refers to the number of times (and whether) an element may or must appear. It is represented by a number, which normally falls into one of the following categories⁷²:

Cardinality notation	Meaning
1	the element is mandatory – and must occur only once
0...n	the element is optional but if present, may occur multiple times
0...1	the element is optional, but if present must occur only once
1...n	the element must occur at least once but may occur multiple times
x or 1...x	the element must occur exactly x number of times (rare), or between 1 and x times (for example, for practicality of storage or display of data; or underlying ontological reasons e.g. a “finger” element might occur between 1 and 10 times for a human)

The “allowed values” column contains listings of the actual allowed value sets, or references to controlled vocabularies, code lists *etc.* Allowed values in [square brackets] are optional.

This way of presenting standards is consciously simplified for ease of reading, comparison and introductory understanding. For example, metadata elements are not divided into simple and composite elements, but rather “flattened” into linear tables, hiding much of the underlying structure. For technical readers wishing to apply the standards, the original schemas must be consulted in all cases.

6.2 Cross-sectoral standards

It is worth noting here that although the DOI system of identifiers and services can be used to identify anything, it is out of scope for this report, since it is primarily an infrastructure standard and its main implementations are in out-of-scope sectors: research datasets⁷³, academic journals⁷⁴ and electronic learning objects⁷⁵.

Two of the identifiers in this report, the ISBN-A (section 6.3.5.) and the EIDR system (section 6.5.2.) are DOI-based; however, the technical specifications and potential business models of DOI registration agencies will not be explored here in detail; this would be more relevant to future deliverables of Linked Heritage WP4.

⁷² Table based on “Mandatory and optional elements” from IFFRO, 2010.

⁷³ The DataCite (<http://www.datacite.org/>) management agency is TIB, a partner on this project.

⁷⁴ Through CrossRef (see <http://www.crossref.org/>)

⁷⁵ Through LON (see <http://www.learningobjectsnetwork.com/>)

6.2.1 ISNI – International Standard Name Identifier

Introduction

The only properly cross-sector standard identifier is the ISNI (International Standard Name Identifier) for identifying the people who create intellectual content of any kind. Some industry sectors have developed their own party identifiers used in specific domains:

Party identifier	Industry sectors	Acronym	Notes	URL
Interested Parties Information	Music	IPI	Developed and primarily used in the music industry but identifies parties involved in all types of creative media	http://www.ipisystem.org/
International Performers Database	Music Film	IPDA		http://www2.ipddb.org/
Virtual International Authority File	Books	VIAF	VIAF is run by libraries, not publishers or collecting societies Since rights are dealt with in-house publishers do not normally need to share authors' identity	http://viaf.org/
Equity names	Film	n/a	Unique stage names are used to identify and distinguish members of the Equity actors' union ("there is ultimately no high-level logical distinction between a title [or a name] and an identifier" – Rust, 1998)	http://www.equity.org.uk/
Picture Licensing Universal System ID	Photography	PLUS ID	The PLUS registry, although in beta stage, already allows registration of parties	http://www.useplus.com/ https://www.plusregistry.org/

There is an obvious similarity between ISNI and the Virtual International Authority File (VIAF)⁷⁶ in the field of book publishing and in fact the partners developing both systems aim to work together, with the VIAF forming a starter database for ISNI registration. The ISNI is a draft international standard at the

⁷⁶ See <http://viaf.org/>

time of writing this report. This standard is planned for publication in 2011 having been in discussion and development for some years.

Scope

The ISNI identifies the persona (a public identity or name used by persons or companies) involved in a creation or transaction. It is thus truly independent of industry sector. Examples of parties in ISNI's scope include:

- natural persons;
- legal entities (such as organisations and corporations);
- fictional characters;
- publicly identified pseudonyms (*i.e.* pen names);
- trade names;
- stage names;
- Brand names.

Construction

ISNI hhhh-hhhh-hhhh-hhh-d(or X)

Fifteen hexadecimal digits, followed by a check digit make up the ISNI. The final check digit may be a decimal 0–9 or “X” (since it is in base 11).

Reference descriptive metadata

The references for ISNI are currently in a test stage, but the ISNI website indicates⁷⁷ they will include the following (it is unclear what cardinality will apply to these elements):

Element	Description
First Name	Personal first name or company name
Last Name	Personal surname or company type
Class	Media type or work type classification
Role	Contributor or creator role

6.2.2 Other areas of cooperation and overlap

There is scope for a great deal of cooperation across sectors, especially within the domain of digitised and born-digital content.

Some products appear to cross sector boundaries:

- Audiobooks – book vs. sound recording;
- Music videos – recorded musical performance vs. AV recording.

Metadata and content vendors have integrated across sectors:

- ROVI – book, music, AV data;
- Getty – stock photographs, AV, music;
- Amazon – retail for books, DVD, CD, music.

⁷⁷ See <http://www.isni.org/db.html> [accessed August 2011]

6.3 Books

6.3.1 The book publishing sector

Overview

The book publishing supply chain in the larger markets (Germany, the UK, France, Spain, Italy and the Netherlands)⁷⁸ is highly integrated, with well-established standards and clearly documented sets of best practices⁷⁹.

Even so, there is a considerable degree of variety, even in the largest markets, as to how the detailed distribution supply chain has developed and is organised today.

Supply chains

The traditional book publishing supply chain typically (though of course not in all markets) follows the basic pattern:

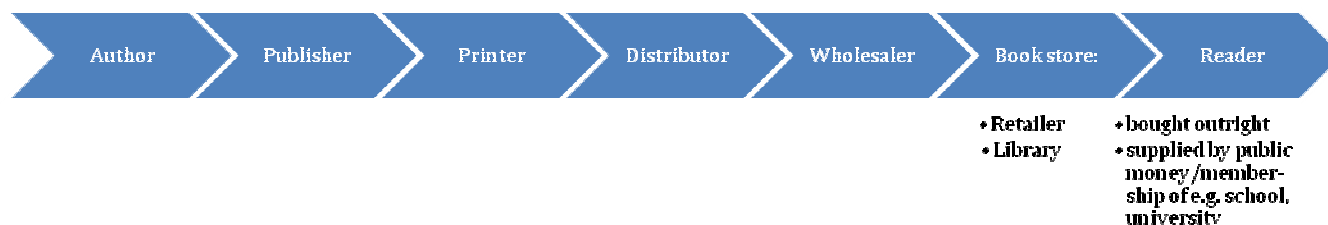


Diagram adapted from International ISBN Agency, 2010.

The supply chain for audiobooks on magnetic or optical media (e.g. tape cassette or CD) has an analogous structure, and is supported by the same standards. However the advent of e-books and e-audio potentially removes several links of the chain, allowing publishers to address end purchasers directly or through just one or two intermediaries, and there may be several new channels for addressing various sub-markets of book buyers. E-book supply chains and markets are in a state of rapid and thoroughgoing change that has caused a crisis situation for standards; publishers are applying the existing ISBN standard inconsistently, or inventing new “standards” with no persistence or uniqueness guarantees (Bide, 2011b).

Standards

Two of the main standards in this sector, the ISBN identifier and the ONIX product information format, are well established and widely used. The ISBN identifies a product, something that can be produced and sold, such as a book or CD of an audiobook. Different editions of a publication, different “bindings” (Simmonds, 1999), or physical formats like hardback, paperback, etc., receive different ISBNs as they will be created, distributed and sold in different ways. The same treatment is officially recommended for different formats of an electronic publication.

⁷⁸ According to the Federation of European Publishers; see <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/619>

⁷⁹ See BISG, 2010 for a detailed supply chain overview, for example: http://www.bisg.org/docs/Roadmap_of_Identifier.pdf

It is interesting to note that the ISBN's scope includes audiobooks, even though these products contain audio recordings which may contain abstractions identified by other standards, for example the ISRC or GRid (see 6.5.4 and 6.5.5).

Due to the confusion surrounding assignment of ISBNs to e-books (despite the official guidance that each file format must receive a new ISBN – see International ISBN Agency, 2010) adopting a kind of “release” identifier like the GRid used for music recordings (see section 6.5) to denote a package of e-book versions of one text has been suggested but was rejected by the industry⁸⁰.

ONIX is used to communicate information about products through the entire supply chain, beginning with ISBN registration, right to the point of delivery to purchasers and sale. It is the standard used by many “books-in-print” services, which keep up-to-date records of books which are available for trading; these services are relevant to Linked Heritage as they often act as substantial aggregators of book metadata, encouraging partners within the supply chain to use applicable standards, and operating a business model that offers a paid metadata service.

6.3.2 ISTC – International Standard Text Code

Introduction

The ISTC (International Standard Text Code) is a new identifier (introduced as an International Standard in 2009) at the work level for books. It has been introduced to facilitate *publishing rights management* and transactions, by identifying the *creative work(s)* embodied in a product (*i.e.* a book with an ISBN), at a higher level of abstraction than the ISBN. The ISTC, with appropriate IT systems in place, “*facilitates the “inheritance” of metadata from work level to manifestation level and can save rekeying*”⁸¹ of bibliographic metadata not relevant to the manifestation in question. It is also intended to assist in tracking rights in works through networks of related (derived, compiled etc.) works (Holdsworth, 2010). Lastly, it can be used both to collocate alternative publications of a work, and disambiguate similarly named titles and authors (International ISTC Agency, 2010).

Scope

The ISTC identifies textual works, defined as “predominantly comprising a combination of words”⁸² that might be manifested in any medium (thus it can identify works manifested in e-books and audiobooks⁸³, grouping together products with different ISBNs in the book industry).

Textual works in scope can include familiar types like novels, or poems, and, interestingly for this report, ISTCs can also identify scripts for audio or audiovisual performances, and so could possibly be linked to the types of AV works identified by ISRCs and ISANs as well; it can also be assigned to song lyrics and so could appear in the packages of intellectual property designated by GRid.

Governance

The ISTC International Agency and its board govern the standard and manage ISTC Registration Agencies in a similar way to ISBN governance (see 8.3.3. below).

Construction

ISTC hhh-hhhh-hhhhhhhh-h

⁸⁰ See, again, International ISBN Agency, 2010

⁸¹ See “How can all the different versions of an e-book be linked together?”
http://www.isbn-international.org/faqs/view/17#q_12

⁸² See International ISTC Agency, 2010

⁸³ See Holdsworth, 2010.

An ISTC consists of 3 fixed-length elements plus a check digit:

Element	Length	Possible values	Meaning
Registration agency	3 digits	0–F	Identification code of the registering agency
Year of registration	4 digits	0–9	YYYY format year of registration
Textual work	8 digits	0 – f	A dumb serial number assigned on accession to the ISTC register, after uniqueness checking
Check digit	1 digit	0 – f	Validation

Reference descriptive metadata for ISTC registration

ONIX for ISTC is the messaging format used to register a work for an ISTC (see EDItEUR, 2009b). It consists of the following core elements:

Data elements	Cardinality	Allowed values	Comments
ISTC	0...1	ISTC	Blank for registration of a new ISTC; completed for confirmation message
ISTC Work Type	1...n	Undefined, Prose, (Song) lyrics, Poetry, Scripts (screen, audio, stage, other e.g. speech)	As with most ONIX formats, the allowed values are in fact expressed as language-independent codes rather than as text
Origination	1	Unspecified, Original, Derived	For linking derived works to their 'parents' (e.g. a translation to the original language work)
Derivation type	1...n (if applicable)	Unspecified, Abridged, Annotated, Compilation, Critical, Excerpt, Expurgated	Allows the relationship(s) between work and its parent work(s) to be described (for derived works only)
Derivation note	0...1		Used to identify a source work in the absence of source ISTC
Source ISTC	0...n	ISTC of parent	
Title type	1...n	Undefined, Original, Uniform, Opening words, Parallel, Other	This controlled list allows typical types of title from various fields of endeavour

Data elements	Cardinality	Allowed values	Comments
Title script	0...n	ISO 15924	
Title text	1		Part of the repeatable “title” composite, including the above 2 lines
Subtitle	0...1		
Edition number	0...1		First, second <i>etc.</i>
Edition statement	0...1		Text describing the particular edition
Work date	0...n		Optional date with a “role” attribute specifying Nominal, First publication or Author activity. Composite with controlled values for calendar and date format.
Contributor role			
Contributor	1...n		
Language of text	1...n	ISO 639-2B)	
Registrant	1		The party seeking the ISTC
Manifestation	0...n		Optional details of one or more concrete manifestations of the work
QueryExistingISTC	0...1	Publisher identifiers, e.g. title, publisher, ISBN	Used to disambiguate in the case of potential duplicate requests
Preferred ISTC	0...1	ISTC	Used to indicate that this work record should be discarded and the preferred ISTC used instead

It should be noted that the above elements are the “work metadata”, extracted from a wider set of elements used in a message choreography for requesting registration, amendments to metadata, deduplication of ISTC records, and cancellation (including all possible replies to these four initiatives).

6.3.3 ISBN – International Standard Book Number

Introduction

ISBN is one of the best known and established product identifiers in cultural industries. “Published as an international standard (ISO 2108) and in wide use since the 1970s it has been adopted in over 160 countries”⁸⁴ and, as of 2007, is integrated within the general retail identifier (GTIN-13)⁸⁵. This means it

⁸⁴ See <http://www.editeur.org/35/ISBN/>

⁸⁵ See http://www.bisq.org/docs/Guide_to_prod_id.pdf

is recognised (although only as a GTIN-13) by systems used by distributors, wholesalers and retailers outside the traditional book trade. It is also used as a standard by libraries.

Scope

“ISBNs are assigned to monographic publications that are available to the public, whether those publications and related products are available on a gratis basis or to purchase. In addition, individual sections (such as chapters) of monographic publications or issues or articles of continuing resources that are made available separately may also use the ISBN as an identifier. With regard to the various media available, it is of no importance in what physical form the content is documented and distributed; however, each product form should be identified separately.”⁸⁶

The following are examples of types of monographic publications in-scope for ISBN87:

- printed books and pamphlets;
- individual chapters or sections of a publication, if these are made available separately;
- Braille publications;
- publications that are not intended by the publisher to be updated regularly or continued indefinitely;
- individual articles or issues of a particular continuing resource (but not the continuing resource in its entirety);
- maps;
- educational/instructional films, videos and transparencies;
- audiobooks on cassette, or CD, or DVD (talking books);
- electronic publications either on physical carriers (such as machine-readable tapes, diskettes, or CD-ROMs) or on the Internet;
- digitised copies of print monographic publications;
- microform publications;
- educational or instructional software;
- mixed media publications (where the principal constituent is text-based).

Any musical scores, or other publications consisting substantially of notated music, are covered by ISMN (see International ISMN Agency, 2008) but the International Standard Music Number is effectively a subset of ISBN.

As mentioned in the introduction this scope has developed over time to reflect a variety of monographic products sold by publishers and booksellers, which do not always clearly fit one category; e.g. maps, transparencies, videos are not “books”. The ISBN’s scope includes e-books. As the International ISBN Agency has directly stated, “the same rules that had been applied to printed books should also apply to e-books”.⁸⁸

Construction

According to the International ISBN Agency⁸⁹ the “ISBN consists of 5 parts with each section being separated by spaces or hyphens. Three of the five elements may be of varying length”.

⁸⁶ ISBN Users’ Manual, new edition pre-publication

⁸⁷ See http://www.isbn-international.org/faqs/view/5#q_3

⁸⁸ See http://www.bisg.org/docs/isbn_agency.pdf

⁸⁹ See http://www.isbn-international.org/faqs/view/5#q_5

Element	Length	Possible values	Meaning (at registration only)
Prefix	3 digits	978 9791–9799	978 or 9791–9799 = ISBN Note that 9790 = ISMN All ISMN values fall within the overall range of numbers used by ISBN, and similarly, ISBN is a subset of the larger GTIN-13 number space
Registration group	1–5 digits	0–9	identifies country, geographical region, or language area of originating publisher
Registrant	1–7 digits	0–9	identifies publisher or imprint
Publication	1–6 digits	0–9	identifies edition and format of a specific product
Check digit	1 digit	0–9	Modulus 10 system, alternate weights of 1 and 3

In the syntax of this report the ISBN appears thus:

ISBN 978-d(dddd)-d(dddddd)-d(ddddd)-d

ISBN 979-d(dddd)-d(dddddd)-d(ddddd)-d

The (brackets) above indicate that the numbers within them are optional, so the element is variable in length. E.g. dd(dd) indicates an element of between 2 and 4 decimal digits. The total length must, of course, remain 13 digits. The first d of the 979 ISBNs can only have the values 1 – 9.

The ISBN-13 is now a subset of the GTIN-13 system for identifying generic “products for sale”⁹⁰, enabling more general retailers internationally to use it within their existing purchasing, inventory and sales reporting systems, and book retailers to use standardised GTIN retail systems.

Reference descriptive metadata

There is a well-defined and documented set of metadata for ISBN registration, available to all registrants. It may be expressed in ONIX for Books version 3.0 (see below for more details) and thus allows publishers to use existing ONIX-compliant systems to register ISBNs.

Data elements	Cardinality	Allowed values	Comments
ISBN	1	GTIN-13 (ISBN subset)	May be assigned by the ISBN Agency, or an agency can assign a range of numbers to a publisher, and the publisher subsequently managed assignment of individual ISBNs to publications.

⁹⁰ See <http://www.gs1uk.org/what-we-do/GS1-standards/Pages/Identification.aspx> for details of GTIN identifiers.

Data elements	Cardinality	Allowed values	Comments
Product form	1		A detailed composite with controlled values for medium and/or format of the product, including: • physical measurements • for digital products, different file formats, licence terms (usage constraints) etc., operating system or other technical requirements
Title	1...n	Title types list	There can be more than one title
Series	0...n		Can include e.g. ISSN for serial
Contributor	1...n	Role and name	May include an ISNI for the contributor(s)
Edition	0...1	Edition types list	
Language(s)	1...n	ISO 639-2/B	
Imprint	1...n	Brand name	May identify the brand with an ISNI
Publisher	1...n	Organisation name	May identify the organisation with an ISNI
Country of publication	1	ISO 3166-1	
Publication date	1	ISO 8601	
ISBN of parent publication	0...n	GTIN-13	Enables registration of chapters of an already registered publication

6.3.4 ISMN – International Standard Music Number

Introduction

The ISMN is fully integrated into ISBN. Being essentially a subset of ISBN, this identifier shares all of the functional characteristics of ISBN above, and uses the 9790 prefix to indicate its ranges of values.

Scope

The ISMN's scope is broadly "notated music"; it complements the scope of ISBN. This includes "other media that are an integral component of a publication of notated music (e.g., a tape recording that is one of the «parts» of a composition"⁹¹, presumably to aid in learning the part) so it covers recordings of works that might be eligible for ISRCs (though the ISRC would identify the underlying recording and the ISMN would identify the product that embodies the recording). Of course musical works

⁹¹ See International ISMN Agency, 2008)

manifested in scores could also be assigned ISWCs (see music section below). The ISMN, however, identifies the manifestation level entity – the printed or digitally represented score in a given product form⁹², as for ISBN above.

Governance

The ISMN is governed by the International ISMN Agency, through regional registration agencies which may also be based within the same organisation as their region's ISBN agencies.

Construction

The construction of ISMN is the same as for ISBN, except for one key difference. There is no registration group element; it uses a subset of the ISBN number range, 9790 and this is followed by a registrant section (publisher and prefix) and a product section, then the check digit.

ISMN 979-0-ddd(dddd)-dd(dddd)-d

Reference descriptive metadata

The ISMN User Manual gives the following minimum registration metadata set, however, only for “illustrative purposes”, and without any cardinality specified for the elements. Although many of fields lack definition or comment in the authoritative documentation, it is clear that they are extremely similar to those specified for ISBN above, and could potentially use many of the same elements, composites and code lists. ONIX for ISBN Registration could be used as a guide for all except 2 of the elements⁹³.

Data elements	Allowed values	Comments
ISMN		Assigned by an ISMN registration agency
Product form	Coding which indicates the medium and/or format of the product	This refers to the “product form”, e.g. hardback, paperback, electronic version, set of leaflet scores <i>etc.</i> , rather than e.g. the parts of the score (contrast “Notated music form” below)
Title		
ISWC	The ISWC of the musical work content, when applicable	Note requirement of “work” identifier in contrast to ISBN's optional inclusion of ISTC for registration
Series title and enumeration		When applicable
Contributor	Contributor role code(s) and contributor name(s)	Composer, lyricist <i>etc.</i>
Edition		Edition number (for editions after the first), type and statement

⁹² “Sheet music in electronic form will be treated the same way as printed material. If there are downloading options for different formats, different ISMNs may be advisable to avoid confusion”, Ibid.

⁹³ See ISMN Newsletter 19, p. 20:

[http://www.cypruslibrary.gov.cy/moec/cl/cl.nsf/all/AEF3F4D9E7377260C225780900393C2D/\\$file/ISMN%20Newsletter%2019_web.pdf?openelement](http://www.cypruslibrary.gov.cy/moec/cl/cl.nsf/all/AEF3F4D9E7377260C225780900393C2D/$file/ISMN%20Newsletter%2019_web.pdf?openelement)

Data elements	Allowed values	Comments
Language(s) of text	ISO 639-2/B	For lyrics
Imprint		The brand name under which the publication is published
Notated music format		The special music format of a work, distinguishing it from other formats of the same work (e.g. Full score, Score and set of parts)
Publisher		The person or organization that owns the imprint at the date of publication
Country of publication	ISO 3166-1	
Publication date		The date of the first publication under this ISMN in the format (YYYY-MM-DD).
Plate number		Especially when plate number constitutes a part of ISMN (when applicable)
ISMN of parent publication	ISMN	ISMN of the parent publication of which this publication is a part, when applicable

6.3.5 ISBN-A – Actionable ISBN

Introduction

The Actionable ISBN, or ISBN-A, is a DOI (Digital Object Identifier⁹⁴) implementation, providing a service layer that resolves an ISBN encoded as a DOI to a product landing page for a given book which can contain metadata about the work, marketing collateral such as reviews and links to other resources, and – using DOI's multiple resolution functionality – multiple instances of links to buy the book (perhaps with links to view or purchase different manifestations of the expressed work). The ISBN agency in Germany⁹⁵ and the Italian ISBN agency⁹⁶ are so far the only suppliers of ISBN-As, in cooperation with mEDRA, the European DOI agency.

Scope

The scope of the ISBN-A is by definition the same as the ISBN's.

Governance

The ISBN-A adds services around the ISBN identifier and thus involves the same governance structures as the ISBN. Therefore, to register an ISBN-A, one needs to contact a DOI registration agency which also registers ISBNs⁹⁷.

⁹⁴ See http://www.doi.org/about_the_doi.html for more on DOI

⁹⁵ German Book Market Standards Agency: www.german-isbn.de/isbn-a

⁹⁶ See <http://www.isbn.it/LISBNA/cos%C3%A8ISBNA.aspx> (Italian language)

⁹⁷ See <http://www.doi.org/factsheets/ISBN-A.html> for this and other practical details.

Construction

The construction of the ISBN code part of the ISBN-A DOI is identical to that of the simple ISBN for the product in question. However, the number is broken into elements (as is normal with a DOI) that separate the registration group and publisher group (see ISBN above for explanation of these groups) from the product identification group:

DOI:10.978.dd(dddddd)/d(ddddd)d

DOI:10.979.dd(dddddd)/d(ddddd)d

The dots and slashes in a DOI are not for display only (as they are for identifiers such as ISBN or ISMN). They are part of the DOI syntax and cannot be omitted.

Element	Length	Possible values	Meaning
Handle System DOI name prefix	2 digits	10	Identifies this code as a DOI for recognition and resolution
ISBN (GS1) Bookland prefix	3 digits	978 or 979	As per ISBN above
ISBN registration group element and publisher prefix	2 to 8 digits	0 – 9	As per ISBN above
Prefix/suffix divider	1 char	/	Standard DOI syntax
ISBN title enumerator and check digit	2 to 7 digits	0 – 9	As per ISBN above

Reference descriptive metadata

These are the same as for ISBN above, with the caveat that the ISBN itself must be registered with a Registration Agency like every other ISBN (see *Governance* above).

6.3.6 ONIX for Books

Introduction

ONIX for Books is the preeminent standard for communication of product metadata between book publishers of all types (trade, educational, academic and STML – “scientific, technical, medial and legal”) and their supply chain partners (data aggregators, logistics organisations, retailers, libraries).

The standard is defined by an XML schema with a highly granular set of data elements, an extensive set of controlled vocabularies, normative documentation and implementation guidelines.

EDItEUR and a number of other trade bodies also produce best practice documents and compliance and certification schemes to ensure the data files exchanged across the supply chain are highly interoperable.

Note that ONIX is a *communication* format, and does not specify any structural details of the database at either end of a message transaction. While a data provider would ideally require a highly normalised data structure to make data management more efficient and to ensure internal consistency, a data recipient may not need to manage the data extensively, and may want to use a highly denormalised data structure for performance reasons. Because there is no consensus across the supply chain about the ideal data structure, ONIX data is serialised. ONIX is not limited to the book industry business

model or the supply chain arrangements of any specific country. It is flexible, and has been deliberately designed for global use.

The latest ONIX for Books is version 3.0. Other versions are still in use, as noted below.

Scope

ONIX for Books covers all products commonly produced and sold through the book publishing industry, including:

- printed books;
- e-books;
- audiobooks;
- maps and charts;
- educational software
- trade only items (e.g. POS, multipacks);
- printed diaries, calendars *etc.*;
- pictures and posters;
- merchandise and promotional materials;
- printed music.

Note that this includes all products that might be identified with an ISBN as well as many extra types (e.g. stationery items like diaries are not strictly eligible for an ISBN); the scope is much broader than that of ISBN. For this reason ONIX codelist⁹⁸ 5 contains a set of standard identifiers including ISBN-10, ISBN-13, ISMN, ISMN-13, GTIN-13, GTIN-14, DOI, LCCN, OCLC number that may be used to identify the subject of an ONIX product record.⁹⁹

Construction

An ONIX 3.0 message (remember that ONIX is a communication format) consists of a header followed by one or more product records. The header contains metadata about the message itself, for example information about the sender and intended recipient(s) and a timestamp.

Message part	Notes
<?xml version="1.0"?>	XML declaration
<ONIXMessage release="3.0">	ONIX declaration and version number
<Header>...</Header>	
<Product>...</Product>	Product record #1
<Product>...</Product>	Product record #2
<Product>...</Product>	Product record #3
</ONIXMessage>	

Outline structure of an ONIX message (from Editeur, 2009a).

⁹⁸ See Appendix 1 – Glossary of terms

⁹⁹ See <http://www.editeur.org/ONIX/book/codellists/current.html>

The body of the message consists of an unlimited number of product records, each relating to a single 'product'. Generally, this can be taken to mean a tradable product (*i.e.* a product for retail sale) such as a book or e-book, but there are exceptions:

- a product record can also be used to describe an individual item which is sold only as part of a set of several items;
- a set of items which are only sold separately;
- a piece of promotional material which is offered to retailers but which is not itself for sale;
- a trade pack intended to be broken up by a retailer for sale to consumers as individual items.

The product record begins with a few elements of record metadata (or 'housekeeping'), among which a record identifier and a coded notification type are mandatory. A product identifier is also required, often in the form of a GTIN-13 (Global Trade Identification Number) – usually this is an ISBN-13, unless the record refers to a non-book item whose GTIN comes from a different source. Other forms of product identifier can also be sent.

The rest of the record is made up of six blocks:

Message part	Description	Notes
<Product>		
Record metadata	Group P.1	Provenance of data
Product numbers	Group P.2	ISBN <i>etc.</i>
<DescriptiveDetail>...</DescriptiveDetail>	Block 1	Main bibliographic metadata that describes the nature of the product
<CollateralDetail>...</CollateralDetail>	Block 2	This enriched metadata could enhance Europeana's discovery environment
<ContentDetail>...</ContentDetail>	Block 3	Not widely used
<PublishingDetail>...</PublishingDetail>	Block 4	Territoriality – retail availability in certain countries?
<RelatedMaterial>...</RelatedMaterial>	Block 5	"links" take the form of references to other works and products – could be converted to Linked Data
<ProductSupply>...</ProductSupply>	Block 6	Could be used for localisation of Europeana search results?
</Product>		

Most of these blocks can occur once only. The last is repeatable, once for each different market in which the product is available. In a 'full' ONIX product record, Blocks 1 and 4 and at least one occurrence of Block 6 are expected, in addition to Groups P.1 and P.2. In other words, a full record should at least identify and describe the product, identify the publisher, and provide supply detail for one or more markets. In practice, it is also unusual for a full ONIX record to be sent without Block 2, since communicating rich collateral content is fundamental to most ONIX for Books use cases.

In a 'block update' record – which may only occur after full record has been exchanged between sender and receiver – Groups P.1 and P.2 must always appear, but Blocks 1 to 6 are all optional. If only Block 2 is sent, then only the recipient's data that was derived from Block 2 in the earlier full message needs updating, and any data derived from Blocks 1 or 3 to 6 should be retained unchanged. Such block updates are a way of greatly reducing the volume of data that needs to be exchanged between sender and receiver through the lifetime of a product.

Block 1 of the product record carries most of the detailed description of the form and content of the product:

- Product form (hardback, paperback, e-book *etc.*, physical dimensions for physical products, DRM information for e-books);
- Product parts (for multiple-item bundled products);
- Series and set titles;
- Product title;
- Contributors (can be multiple contributors, with various roles such as author, illustrator, translator *etc.*);
- Conference details (for conference proceedings);
- Edition (first, second, third, *etc.*, but also abridged, large print, facsimile *etc.*);
- Language (of text, and of original text if translated);
- Extents (number of pages, file size of e-book *etc.*);
- Illustrations and ancillary content (details of ancillary content);
- Subject (genre/subject of book);
- Audience (intended readership).

Block 2 of the product record carries information relating to various kinds of marketing collateral, either carried in the ONIX record or referenced elsewhere, by web links or otherwise.

There are important distinctions between:

- text which is carried as part of the ONIX record (and which by implication is available to be used directly by the receiver of the record);
- content which is not carried in the ONIX record, but which is offered for linking or download by another supply chain party to support sales, and is available for use by the receiver, subject to any stated terms;
- content which belongs to a third party and is cited for look-up only, and which remains subject to third party copyright.

The collateral material commonly includes a description of the product (intended for use in an online store), an image of the product (for example, a picture of the front cover of a book), text samples, review quotes, biographical notes relating to the contributors, and more rarely includes reading guides, images of the author, audio or video clips, links to full reviews or feature material available online.

Block 3 of the product record allows tables of contents to be carried in a detailed and fully structured form. However, in most applications it has been found sufficient to send tables of contents as a text in Block 2.

Block 4 of the product record covers information about the imprint (brand) and publisher, about the global publishing status (including the publication date), and about the sales rights (rights concerning where the product can be sold, not necessarily the same as the rights the publisher holds over the underlying work).

Block 5 allows links between the product and other products, and between the product and other works to be specified. Thus a product record may include identifiers for other products that contain the

same content, identifiers for other related products (with different content), an identifier for the work of which the product is a manifestation, and identifiers for related works.

Block 6 of the product record covers supply and availability detail for an implicitly or explicitly defined 'market'. The complete block is repeatable for each different market where the product is available. Each repeat includes:

- the geographical description of the market;
- the status and availability of the product in the market;
- a range of potential suppliers for the product, and their prices.

Other versions and variants

ONIX 2.1 is the previous version of ONIX for Books, and remains the most commonly implemented. Much of the overall message structure is very similar, but ONIX 2.1 is missing some elements required for full description of digital products. Its handling of sets, series and multi-item products, and of marketing collateral material is less sophisticated. Specification of complex international sales rights, markets and distribution arrangements is less certain.

There are also other related variants of ONIX for specific purposes and supply chains, along with several piloted versions of particular interest for Linked Heritage, namely ONIX for Multimedia and ONIX for Still Images. Since the latter remain in draft form, there are mentioned here purely for completeness. Sample ONIX messages can be found in Appendix 3.

6.4 Music and Sound

6.4.1 Music industry supply chains

Music industry standards are well developed, because of the need to manage the copyrights and performance rights of many interrelated parties. The high level of metadata schema articulation reflects the fact that two complex supply chains are in play¹⁰⁰. One route through the industry controls rights in musical *works*¹⁰¹; the *expression* of those works in performances (events) and recordings (manifestations) forms the other, dependent on the first, supplying *recorded* music from recording artist to audience; whether digitally (online), on magnetic storage media (tape) or optical discs (CD). The two aspects are intimately intertwined; the originator of a musical work (e.g. composer, songwriter) can also be the performer (e.g. conductor, singer) so the two distinct sides of the supply chain often involve the same actors (and possibly simultaneously –see 6.4.2. on improvised musical works). The following diagrams will assist in exploring the details of both aspects of the industry, leading to a synthetic view highlighting the standards used.

In the simplified and generalised “industry” diagram below, the complete situation is depicted in terms of activities directly adding value, with the two rights chains beginning with “Artists” (blue flow directions) and “Composers” (green flow) distinct but interlinked.

¹⁰⁰ See <http://www.generator.org.uk/sites/generator.org.uk/files/The%2BMusic%2BUniverse.pdf> for a diagram showing the clear distinction between the two sides of the supply chain, in this case represented by value (cash) flows

¹⁰¹ Which are also manifested in published scores and lyrics

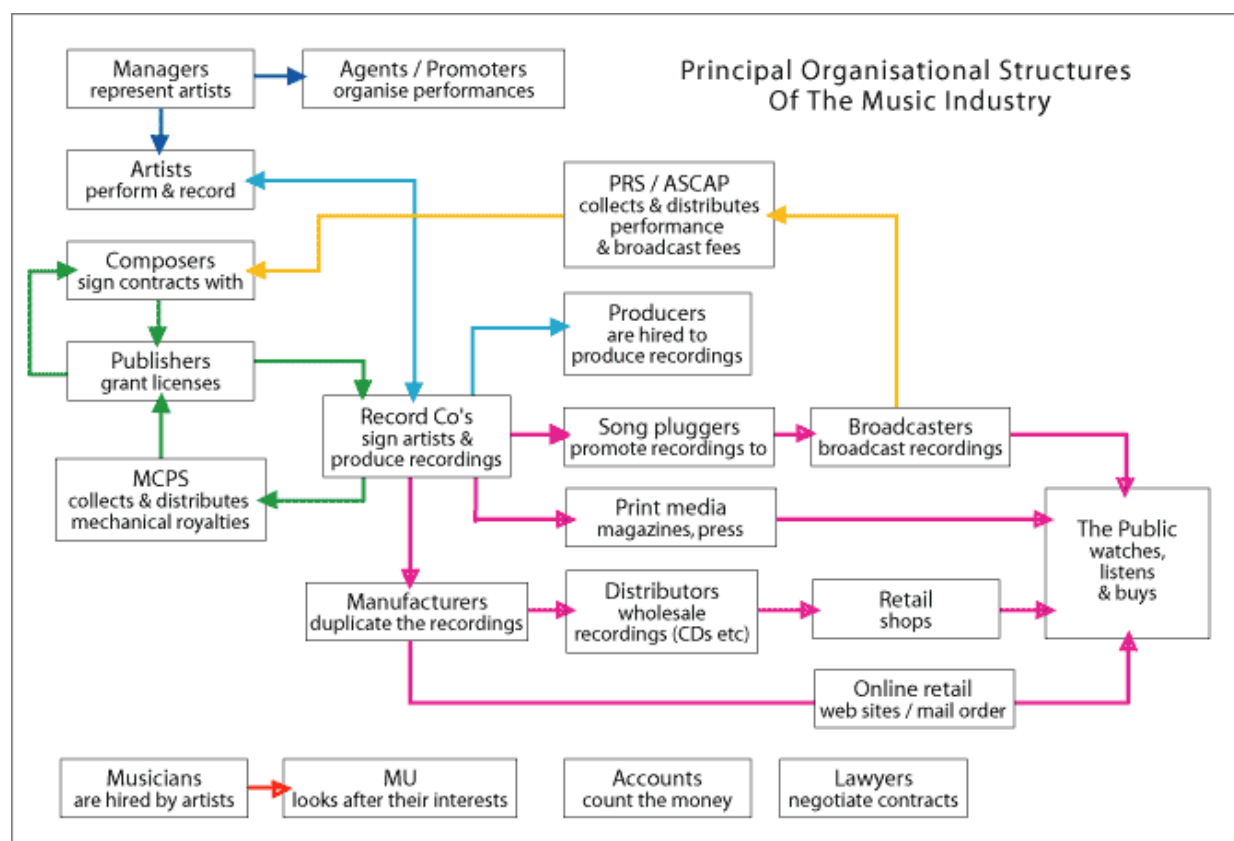


Diagram by Dustry, I. (1999). http://www.planetofunes.com/industry/industry_structure.htm

Musical works rights supply chain

On completion of a new composition (song, orchestral work, etc.) the composer intending to create or licence recordings registers a copy of the notation for the work with an authors' society (Rust, 1996). This enables the management of rights in the as well as furnishing an identifier (the ISWC) for use in the rest of both supply chains. The contributors themselves receive an ID number (IPI – "Interested Parties Information", similar to the ISNI; see section 6.2.1.) and a role code consisting of one of the following¹⁰²:

Code	Meaning
A	Author
C	Composer
CA	Composer/author
AR	Arranger
E	Publisher (French; <i>Editeur</i>)
AM	Administrator
SE	Sub-publisher

¹⁰² See: http://en.wikipedia.org/wiki/Interested_Parties_Information

There is then a linkage midway within the recording industry where artists, record labels, distributors and retailers make deals (or exploit statutory licences) to record and release versions of the works registered in the rights supply chain. The two diagrams below, based on the UK music publishing and recording industries (but applicable elsewhere by analogy), illustrate the detailed operation of the two parallel supply and value chains, their complexity and essential dependence on rights and licencing:

MAP 2 : UK Music Publishing Company

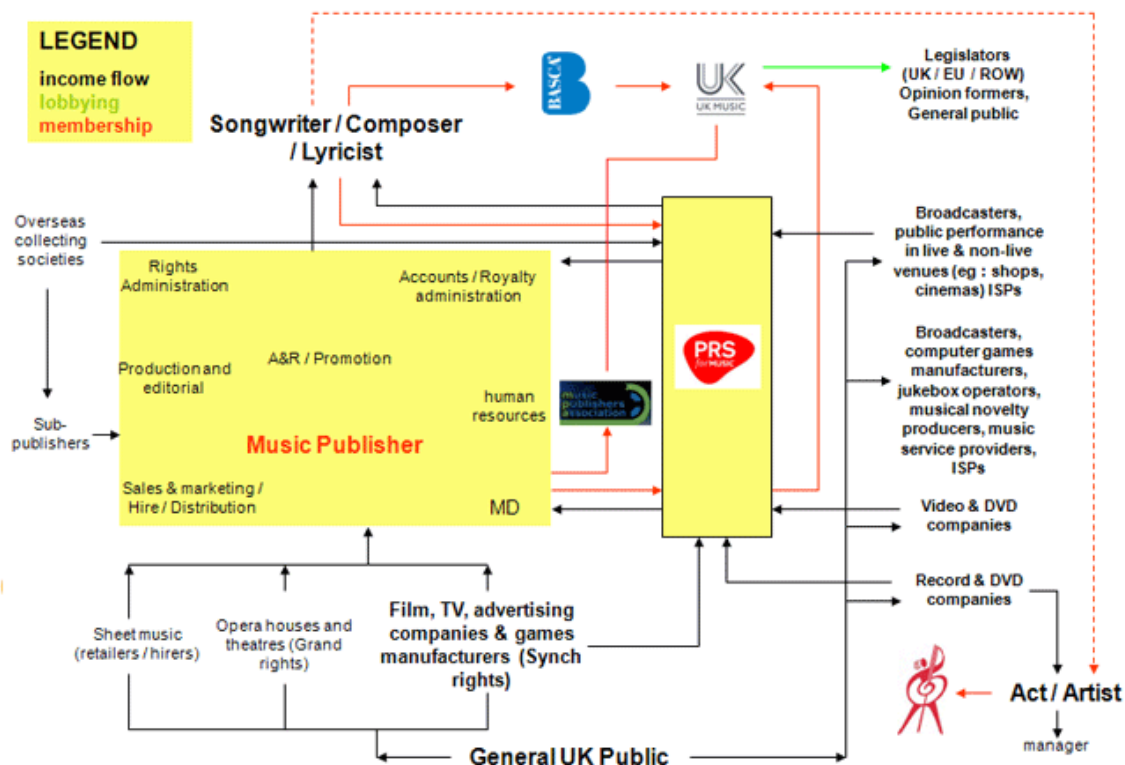


Diagram by BPI & Creative Cultural Skills, 2009

At the performance stage, the ISWC and IPIs serve to identify works and parties for payment of royalties when the work is performed, both in live events, and (or) at the end of the recording industry supply chain from sale or broadcast of recordings.

Recorded music supply chain

In this linked but distinct supply chain, recordings are created, released, fixed in some discrete physical package like a CD or online digital form, distributed and sold to the general public. It is important to note that several other streams of mechanical “performance” (e.g. broadcast, webcast, stream) of recorded works run in parallel to the live performances in the previous diagram; these provide another link between the two chains, and highlight the central importance of Performance Rights and Mechanical Rights Organisations to this business.

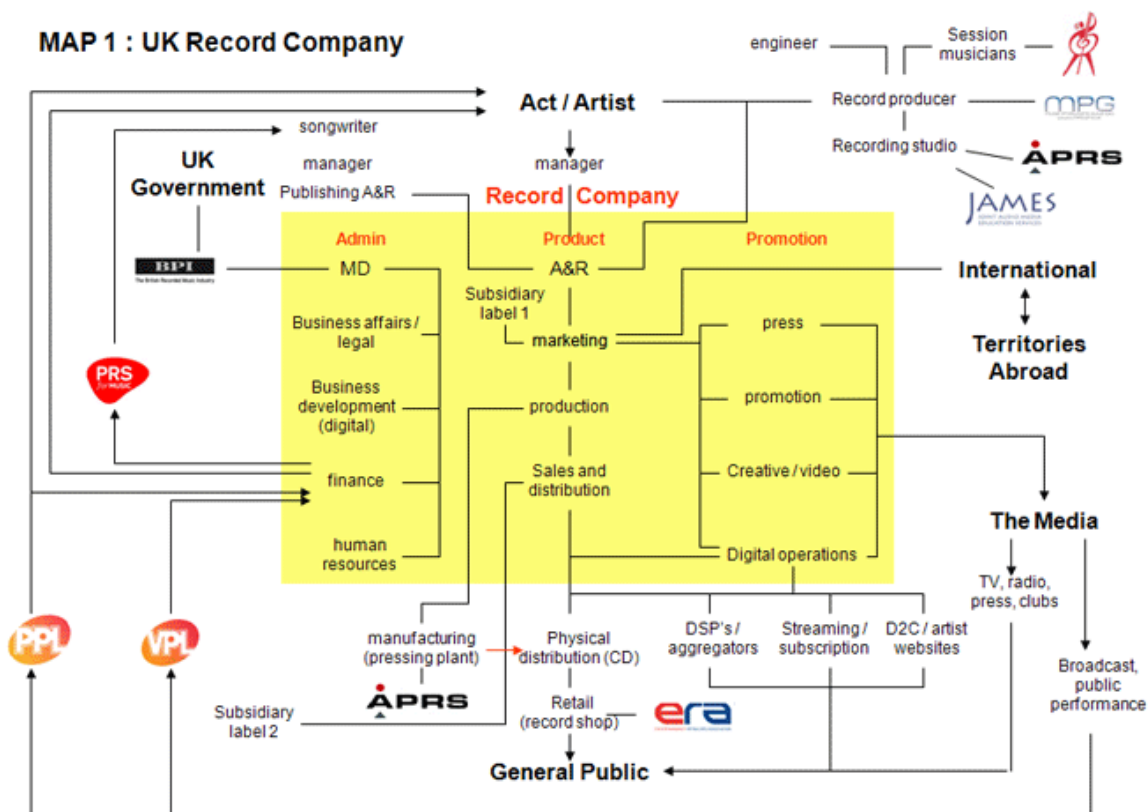


Diagram by BPI & Creative Cultural Skills, 2009

In this second part of the overall music industry ecosystem, musical works are expressed in performances, recorded and manifested in products. These are grouped together as physical products or electronic “releases”, and traded. This part of the ecosystem has been adapted to the digital content era, at least in principle and features some well-developed standards such as the ISRC and GRid.

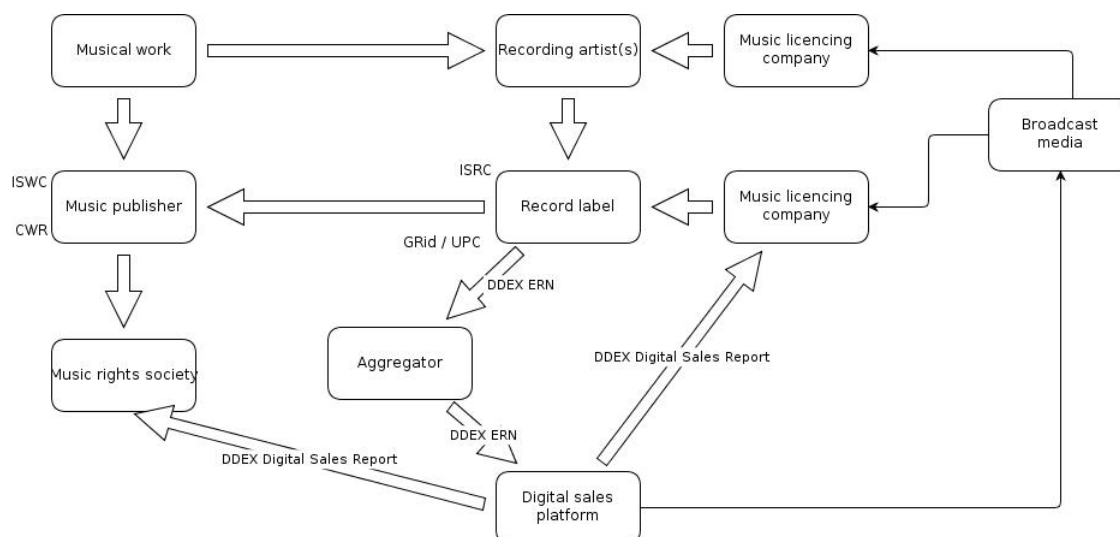


Diagram based on an original drawing by Niels Rump (2011).

The final diagram above summarises the most important relationships and the points where identifiers are assigned and standardised electronic messages are exchanged. The acronyms used will be explained in the sections below, and appendix 2.

The diagram indicates where a standard DDEX message is used. For the transactions and information flows where no DDEX message is used, some sort of proprietary format will apply.

6.4.2 ISWC – International Standard Musical Work Code

The ISWC identifies abstract musical works at the first stage of the artistic production process. It is also used within the music industry supply chain for identification of works “abstracted from recordings”¹⁰³, royalty computation, and of course to protect the composer’s legal and moral rights. Although an ISWC may be assigned before performance, the work may be abstracted from the performance at a later time, especially in the case of extemporaneous forms such as folk, jazz and free-form music.

Scope

ISWC was originally intended as an abstraction and generalisation from ISMN to allow consolidation of many rights societies’ proprietary identifier formats (see Hill, 1997, p. 13.) and now applies to the following types of work¹⁰⁴:

- Dramatico-musical work;
- Musical arrangement of a work;
- Adaptation of the lyrics of a work;
- Translation of the lyrics of a work;
- Instrumental version of a work with associated lyrics;
- Cadenza;
- Recognised excerpt of a work;
- Medley;
- Potpourri;
- Revision of a work;
- Individual musical work (cue) contained in the soundtrack of an audiovisual work.

The final item on this list highlights the use of the ISWC in “cue sheets”, listings of musical works incorporated into an audiovisual work¹⁰⁵, which are used to collect royalties (see above music industry maps and the section on AV).

The inclusion of several “versions” and “adaptations” in the list above (as well as the “versions” element in the reference metadata, which should be another ISWC) demonstrates some complexity in describing the work with its related works¹⁰⁶. This is a similar approach to that of ISTC (see section 8.3.2).

According to Hill (1997, p. 13), the ISWC was intended for extension to other types of works, e.g. literary works and potentially joint management with ISAN (ibid., p. 14.). Use of the ISWC for other works could cause confusion, e.g. with the ISTC (see books section).

The specification for ISWC also makes clear that even non-copyright protected works can receive an ISWC.

Governance

The International ISWC Agency is an ISO-appointed agency, managing regional agencies in the usual manner. The International Agency is based within CISAC (International Confederation of Societies of Authors and Composers)¹⁰⁷. Local agencies correspond to the areas covered by authors’ societies (see link in footnote 100). This highlights the role of ISWC in royalties and rights management.

Construction

¹⁰³ C.f. Rust, 2005a.

¹⁰⁴ List taken from: <http://www.iswc.org/en/creators.html>

¹⁰⁵ For more on cue sheets see: <http://www.ascap.com/playback/2005/winter/features/cuesheets.aspx>

¹⁰⁶ See <http://www.iswc.org/en/faq.html> for more on versions, excerpts and adaptations receiving new ISWCs

¹⁰⁷ See <http://www.iswc.org/en/contact.html>

The ISWC is a simple serial number:

T-ddddddddd-d

A letter “T” denotes the number as an ISWC, and the body of the identifier consists of nine decimal digits. The final digit is a check digit.

Reference descriptive metadata

The set of minimum registration data below is taken from the ISWC International Agency website¹⁰⁸, with cardinality added by this report’s author:

Data elements	Cardinality	Allowed values	Comments
Title	1		Description of the work by the composer
Contributors	1...n	IPI numbers IPI role codes	All composers, lyricists, arrangers of the work
Classification	1	CIS standards list	
Versions	1...n	Other ISWCs	For example arrangements

6.4.3 ISRC – International Standard Recording Code

Overview of ISRC

The ISRC is a standard for sound recordings and music video recordings, which is currently undergoing considerable revision. Uniqueness of each identifying number was very important at time of initial specification, but a central registry of works was not desired by the industry. The highly-distributed structure was designed to enable uniqueness with minimal checking at local level. The revision of ISRC currently under way under the aegis of ISO is expected to establish a central registry for codes and make registration conditional on compliance with the specification.

Use of ISRC as a standard by major distributors means that uptake is extremely high across Europe, and can be assumed to be effectively complete coverage of commercially relevant recordings.

Scope of ISRC

The ISRC identifies sound recordings (musical, spoken word and wildlife sounds). This can therefore include expressions of abstract compositions identified by the ISWC (see above). These recordings (in digital form) are the core components of the “releases” identified by the GRid discussed in section The ISRC is “format independent”, so that it does not change on release in any physical recording medium or file format (e.g. CD versus magnetic tape; MP3 versus AAC).

However, different “versions” of a recording should be assigned different ISRCs. The reasons given for assigning a new ISRC are:

- New recording or material change;
- Re-mastering (involving substantial creative input);
- Remixing;
- Re-issue with significantly changed playing time;
- Change from multi-channel to stereo mixing;

¹⁰⁸ See <http://www.iswc.org/en/faq.html>

- New takes of a work during a recording session (if intended for commercial release);
- Reissue of an out-of-copyright and previously unregistered recording;
- Secret tracks.

The definition of “recording” used in the context of ISRC also includes samples or “snippets” of a recording used for preview or promotion of a whole recording, music videos used to promote a track, and audio-visual recordings of musical performances (so for example, a live concert video or a promo video could receive both an ISRC and an ISAN). Other types of audio recording such as interviews with performers are recommended by International ISRC Agency to receive an ISRC if they are intended or considered likely to be commercially exploited at some point.

Construction of ISRC

The ISRC consists of 12 characters, “is alphanumeric, using digits (the ten Arabic numerals 0 – 9) and the 26 letters of the Roman alphabet”¹⁰⁹, divided into four elements:

Element	Length	Possible values	Meaning
Country	2 alphabetic characters	ISO 3166-1-Alpha-2	Country of residence of registrant
Registrant	3 alphanumeric characters	Assigned by the National ISRC Agency	Assigner of the ISRC to the recording
Year of Reference	2 digits	Last 2 digits of year ISRC is allocated: “This allocation will normally take place in the year in which the preparation of the final production Pre-Master for the recording is finalised”	Makes it easier to guarantee uniqueness of ISRCs
Designation	5 decimal digits		Unique (within registrant) code for this recording

The complete construction is:

ISRC aa-aaa-aa-ddddd

It should be noted that there is no check sum digit in the ISRC as it does not consist solely of numbers (it would be possible to create a checksum by assigning numeric values to the letters, but a checksum was not felt to be vital when the identifier was standardised).

Reference descriptive metadata

According to the latest version of the ISRC handbook, “Work has begun on standardising a minimum metadata set. This is being done by the International ISRC Agency, in co-operation with users and the National Agencies” (International ISRC Agency, 2009). The GRid Handbook currently specifies¹¹⁰ the following minimum elements for each ISRC used in a GRid record:

¹⁰⁹ ISRC Handbook, 3rd Edition

¹¹⁰ See http://www.ifpi.org/content/section_resources/grid-handbook.html#1.9.3.ISRC%20Metadata%20outline

Data elements	Cardinality	Allowed values	Comments
ISRC	1	ISRC	
Title	1		Free text field (but should exclude any annotation that is included in Recording version description)
Recording version description	0...1		This serves the same purpose for the ISRC as the equivalent in GRid metadata (see Section 6.4.4 below)
Name of main artist	1...n		
Reference date	1		If the ISRC identifies a sound recording, the ® date (that is the year of first publication) of the recording in the form (P)YYYY. If the ISRC identifies a music video recording, the © date (that is the year of first publication) of the music audio-visual recording in the form (C)YYYY
Duration	1	ISO 8601:2004	
Medium	1	Sound, AV	

The unpublished proposal for a revised ISRC ¹¹¹ specification, while adopting the same broad format as legacy ISRCs to ensure backward compatibility, uses the country code embedded in the ISRC at registration in a different way and allocates the remainder of the code centrally to ensure data quality and uniqueness.

6.4.4 GRid – Global Release Identifier

Overview of GRid

The GRid is an industry standard, developed by RIAA and IFPI during the MI3P project to create identifiers for digital music, and now governed by the International GRid Authority through the IFPI Secretariat¹¹², with Issuer Codes being governed by the International GRid Registration Agency at RITCO¹¹³.

Scope of GRid

The GRid is a type of identifier unique to the recorded music industry, identifying a *release* which corresponds essentially to the concept of the specification (track listing *etc.*) of one of: a (record) album, single or similar collection of related music recordings, plus any extra content (generally it will be images, text, perhaps AV and metadata all covered by copyright). It can thus identify:

- Single tracks;

¹¹¹ Candidate revised working draft of ISO/DIS 3901(E). [unpublished]

¹¹² See http://www.ifpi.org/content/section_resources/grid.html

¹¹³ See section 11, “Administration of the GRid” in http://www.ifpi.org/content/library/GRid_Handbook_V2_0_final.pdf

- Singles;
- Extended plays;
- Albums;
- Multimedia packages.

The description of the “Release” concept as given in the GRid Handbook reads as follows:

“A record company decides that it wishes to market a specific group of sound recordings available to the public. It may also want to bundle these sound recordings with some other Digital Resources perhaps some textual notes, some still images and a music video. This “bundle” of Digital Resources is the Release which can be identified with a GRid.

The Release is therefore an abstract concept; it can be referred to entirely separately from any particular representation of those Digital Resources. It is also conceptually different from a Product. The same Release may be made available for sale, for example, through different channels, may be consumed by the user in different ways (e.g. streaming, downloading), and provided in different file formats, at different price points.

Each permutation of product attributes might reasonably be seen as a different *Product*. Assuming the Digital Resources remain the same, each Product embodies the same Release.

So, although a Release is the basis of a Product, it is not the Product.”

This collection level identification is required by the way that digital distribution and retail makes possible the unbundling of recordings from their historically mechanical, magnetic or optical media release context, and the need to provide continuity with the legacy systems and working practices underlying this context. There is a similarity with the “digital item” concept used by MPEG-21 (see Burnett, I., 2006).

Construction

The GRid consists of four elements, comprising alphanumeric characters (numbers 0–9 and letters A–Z).

A1-CCCCC-CCCCCCCCC-C

Element	Length	Possible values	Meaning
identifier scheme	2 characters	A1	“A1” identifies a GRid
issuer code element	5 characters	0 – Z	unique code allocated to each Issuer of GRids by the Registration Agency
release number element	10 characters	0 – Z	unique number allocated by the Issuer to each release which they control. Issuer to ensure that each release number element that they use is unique within their own Issuer Code (and that each Release is assigned only one release number element)
Check Character	1 character	0 – Z	

Reference descriptive metadata

The information below is taken from the GRid handbook (IFPI, 2007b). Note the nested elements within “Resource Group” – this element is the actual listing of resources included in the release. Resource Group Identifiers can themselves be included within a Sequenced Resource Group, to build up structured hierarchies of resources within a message. In the format used by this report it has not been possible to represent the structure of the XML but this can be seen from the diagram below the data elements table.

Data elements	Cardinality?	Allowed values	Comments
GRid	1	GRid	
title	1	free text	Issuer may decide which title to use if there is more than one
main artist	1		
Sequenced Release flag	1	Yes, No	a Release in which the order in which some or all of the Digital Resources is material
list of Unsequenced Resources.	0...1	For each sound recording or music audio-visual recording, the ISRC For each other Digital Resource, an alternative identifier (ideally ISO) in the form namespace:identifier, or a free text description of the Digital Resource	
Resource Group (the four bulleted elements below are nested sub-elements of this one)	0 (1...n. if Sequenced Release flag= Yes)		These elements indicate the tops of hierarchies of grouped resources
Resource Group Identifier	1	Free text	
Sequenced Resource Group flag	1	Yes / No	Tells the parser or reader if the ordering of this list matters

Data elements	Cardinality?	Allowed values	Comments
Unsequenced resource in Resource Group	0 (1...n if Sequenced Release flag=No)	ISRC:IDENTIFIER NAMESPACE:IDENTIFIER Free text ID (e.g. semi-unique string detailing title, composer <i>etc.</i>) Another resource group ID	
Sequenced resource in Resource Group	0 (1...n if Sequenced Release flag=Yes)	ISRC:IDENTIFIER NAMESPACE:IDENTIFIER Free text ID (e.g. semi-unique string detailing title, composer <i>etc.</i>) Another resource group ID	The order of the resources will be determined by the actual order of the list of these elements
Resource Context	0-1	GRid:NUMBER NAMESPACE:IDENTIFIER Free text	only permitted when a Release is of a single sound recording used to identify the context (normally an album) from which a track is being released as a single-recording Release
Release Version Description	0...1	Free text	
GRid Issuer	1		

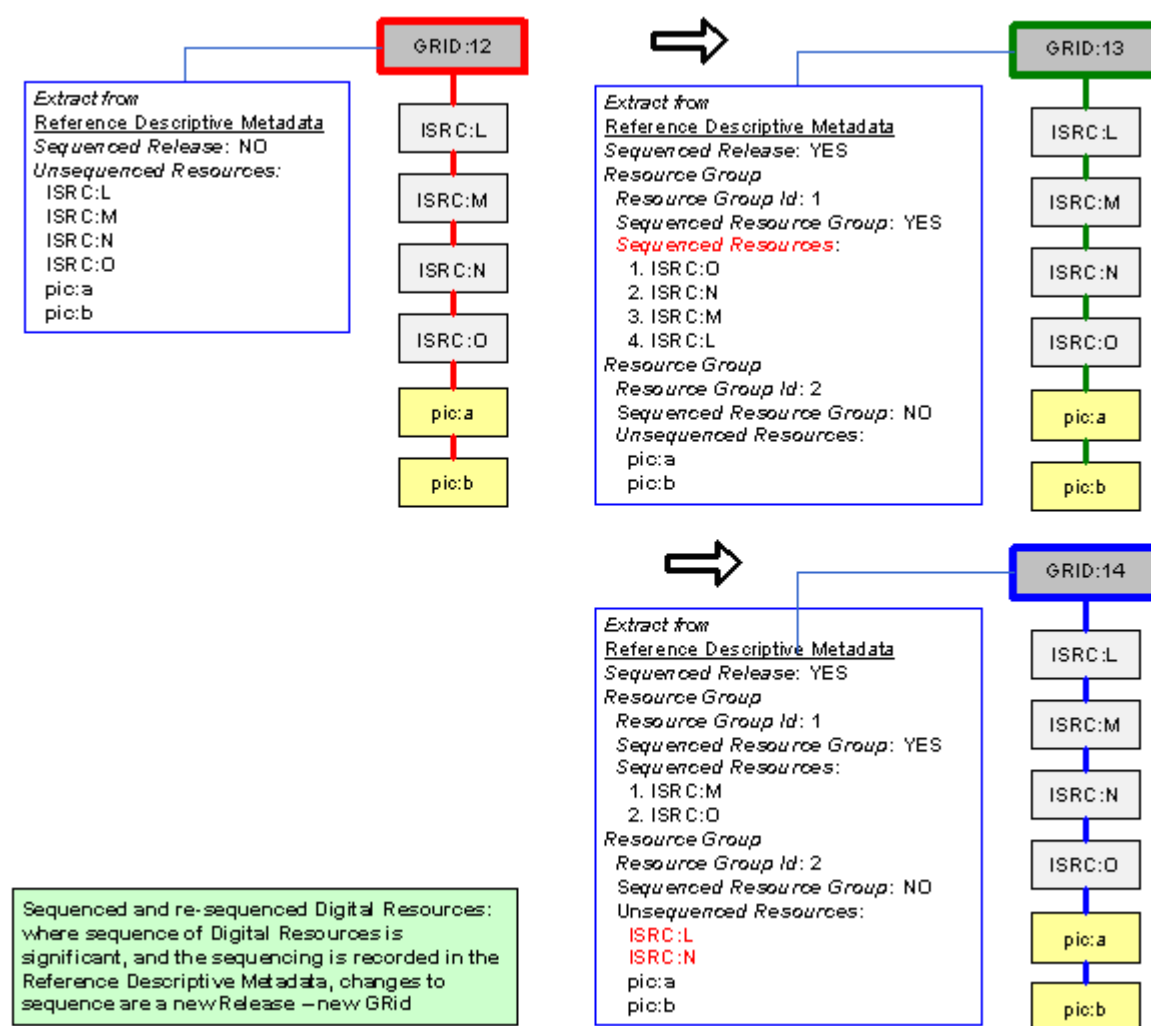


Diagram illustrating structure of 3 possible GRid registrations with sequenced and unsequenced Resource Groups¹¹⁴

6.4.5 DDEX Release Notification

Introduction

DDEX (Digital Data Exchange LLC.) is a membership organisation in the recorded music industry that maintains and governs a “suite of messages that give a uniform mechanism to enable Release Creators (usually record companies) to inform DSPs [Digital Service Providers] about Releases – whether they are newly created or “back catalogue” – that are available for distribution to consumers.”¹¹⁵

Although there are 19 standards managed by DDEX, the one of interest to Linked Heritage is the Release Notification message, whose primary entity is the release identified by the above described GRid. The standard for the message clarifies the relationships between musical works, recordings, releases and products in a commercial context and so provides the justification for the abstraction identified by a GRid (see the glossary in Appendix 1 for full definitions).

Although the entities described are different (abstract groups of resources for GRid and DDEX, manifestations or products for ISBN and ONIX) there is a similarity with the situation in book

¹¹⁴ From: http://www.ifpi.org/content/section_resources/grid-handbook.html#1.12.0.A.5.%20Sequenced%20resources%20%28see%20Figure%20A.5.%29outline

¹¹⁵ See DDEX ERN, available from <http://www.ddex.net/documents.html>

publishing, in that the descriptive metadata standard comprises a set of messages tailored to business needs at various points in the supply chain. These are identified in the final diagram of section 8.5.1, “Music industry supply chains”.

Scope

The scope of DDEX Release Notification includes the releases identified by GRid – but since DDEX provides a complete toolkit for communication along the recorded music supply line, it also includes:

- Musical works and recordings;
- Audiovisual works and recordings;
- Parties;
- Rights shares;
- Cue sheets;
- Collections (e.g. series of audiovisual works);
- Deals;
- Structure (“chapters”) of a work (e.g. movements of a concerto);
- Exploitation of releases (*i.e.* payment methods and structures, allowed uses of a recording, distribution channels and devices used to access recordings).

Again, see the DDEX ERN document (link in footnote 107) for full details.

Note that this scope overlaps considerably with that of other schemes and identifiers, notably for the audiovisual industry, and for musical works (as opposed to recordings). The relevant standards (such as ISWC, ISAN and ISRC) are referenced at the appropriate points.

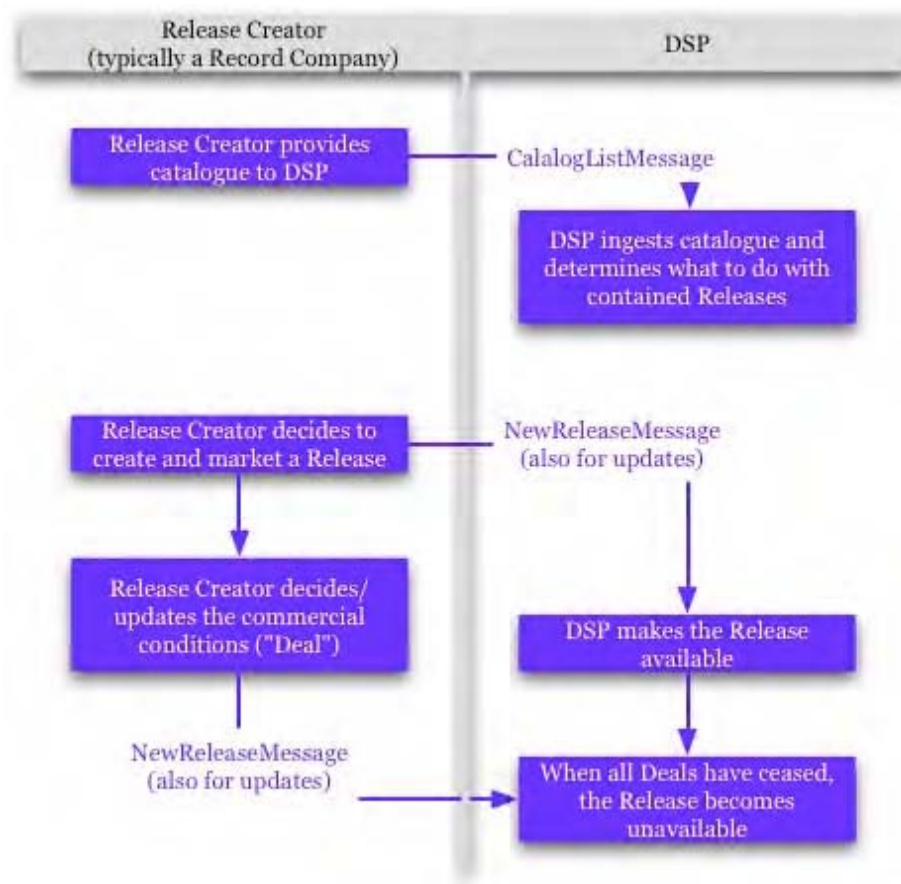
Governance

The membership and board of DDEX, the not-for-profit body that owns the standards, manage all developments.

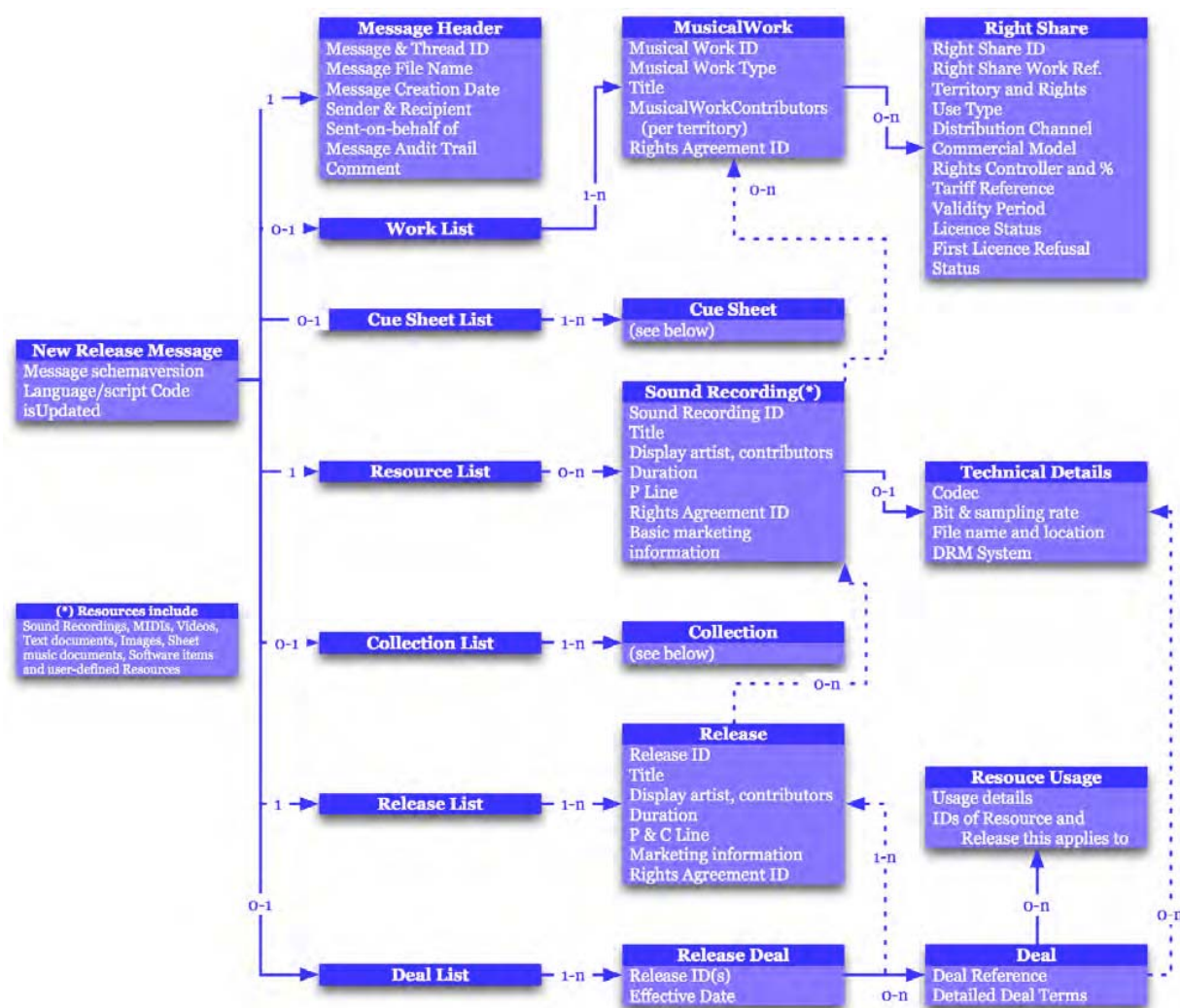
Construction

The DDEX standards contain message structures (including identifiers and metadata), message choreographies (the anticipated sequences of communications initiated and responded through the provided message structures), and guidance as to which elements may be contractually (as opposed to technically) mandatory).

It is beyond the space limitations of this report to overview the entire suite but two examples are reproduced here to give readers an understanding of what is available.



Message Choreography for the DDEX Release Notification standard



Message structure overview for a new release in DDEX

The new release message in DDEX is the part of most potential interest to Europeana because it contains identifiers and descriptions needed for discovery of resources and products, as well as rights and ownership information that may be pertinent to the legal-technical solution.

Other versions and related data models

The DDEX suite of messages is based on the index model as described in 5.4. Other DDEX standards identify and describe parties, deals, musical works licences, sales figures, data exchange rules and mistake correction, digital signatures and the data dictionary¹¹⁶ used in all the metadata standards.

6.5 Film and Television

6.5.1 The Audiovisual Sector

Overview

Commercial audiovisual media can be broken down into two main categories: filmed entertainment (cinema films, “movies”) and television programmes (including series and documentaries). Television news and advertising are out of scope of this report for the same reasons as newspapers and design

¹¹⁶ See Appendix 1 – Glossary of Terms for definition; effectively an ontology or conceptual model

industries. The two top categories will be dealt with here in turn, although the same identifiers and metadata tend to apply to both.

Funding of the film industries through advance payment and risk sharing by distributors leads directly to the granting of exclusivity, territorial and language rights in individual films. Even a high level of public subsidy in the EU does not change this basic situation.

The European cinema film industry is small in global terms, and dominated by five large producers: “Germany, the United Kingdom, Spain, France and Italy together account for 67% of European production. Yet the total of their respective productions in 2000 amounted to only 71% of American production” (IMCA, 2002). The EU is internally characterised by a diverse and fragmented selection of national film sectors, often in high proportion (42-60%) state- or EU-funded, and a large, more coherent Europe-wide distribution and exhibition sector that is commercially dependent on the United States’ film producers: “the market share [in Europe] of the American films represents 66% [...] whereas the group of European films represents only 4.51% of the American market for the same year” according to the European Audiovisual Observatory’s Focus 2002 World Film Market Trends study (quoted in IMCA, 2002).

Further data from a 2002 European Parliament research report (below) show how completely USA funding dominates European film funding:

Country	Company	Quota(%)	Foreign capital (%)	Origin
Germany	UIP	21.7	100 (UIP)	USA
Spain	Buenavista	16.4	100 (Buenavista)	USA
France	Gaumont Buenavista	18.8	50 (Buenavista)	USA
Sweden	Fox Films	27	100 (Fox)	USA
Ireland	U.I.P	25	100 (U.I.P.)	USA

Leaders in cinema film distribution (selected countries):

** Figures from 1999 – Source: School of Public Communication.
University of Navarra; quoted in STOA, 2002.*

Given this state of play it is reasonable to assume that any industry standardisation is driven from the USA, and in practical terms this means the “Big Six” distributors with the greatest market share (not necessarily the most films):

Rank	Distributor	Movies	2011 Gross (as of August 2011) / \$ millions	Share
1	Paramount Pictures	15	1,325	21.28%
2	Warner Bros.	26	1,167	18.75%
3	Universal	12	794	12.75%
4	Sony Pictures	18	722	11.60%

Rank	Distributor	Movies	2011 Gross (as of August 2011) / \$ millions	Share
5	Walt Disney Pictures	12	719	11.55%
6	20th Century Fox	11	559	8.99%

Source: Nash Information Services, LLC. (2011).

All of these are founder members of MovieLabs, which describes itself as the “research and development joint venture started by the six major motion picture studios”¹¹⁷, and most are listed as promoters of EIDR¹¹⁸ (see 6.5.2. below for technical details), the main filmed entertainment industry identifier and registry standard.

Supply chains

Bloore (2009) describes the European film sector value chain as on a par with American “indie” (independent) film production. His description synthesises and builds on many models of the industry,



including the American major studio model, where cinema production, distribution and exhibition are highly integrated within each large company; his model describes the structure of more fragmented European sub-sectors in detail, and integrates the value chain ending in cinema exhibition with the “long tail” value chain of post-release purchases on DVD, film-on-demand etc.

This auxiliary chain based on rights and licensing (or “library long tail” as Bloore calls it) has great importance for European films where the “big hit” on initial release may enjoy a long shelf life on DVD or digital. The time of commercial films also leads to a “windowing” phenomenon where films are released in different formats is phased, and release to specific territories embargoed depending on the progress of releasing in other territories, media and formats.

Library long tail
• Distributor- reissues during first license • Producer- sells new license after reversion • Producer- remake / sequel (back to start)

Rights clearance and licensing

Various different models exist for licensing and distribution in the “long tail”, due to different approaches to the rights in recordings. For films, the rights are normally assigned to the producer, so it is relatively simple to sub-licence them (though territoriality remains an important problem).

For Video on Demand (VoD) access is mostly through set-top boxes rather than through the Internet, which may be a discouraging sign for Europeana.

Television rights clearance is much more complex, since the current system was based on a model in which a recording was assumed to be broadcast once only, and frameworks vary across the EU. The

¹¹⁷ See Motion Picture Laboratories, Inc.’s “about us”: <http://www.movielabs.com/AboutUs/index.html>

¹¹⁸ See <http://eidr.org/membership/>

so-called “Nordic model”¹¹⁹ of TV rights clearance assumes that all rightsholders would have opted into a blanket license and their remuneration is left to rightsholder representation organisations.

Standards

As might be expected from the above consideration, the “six major motion picture studios” together fund the MovieLabs research venture¹²⁰ which have led to the creation of EIDR, one of the two main identification standards for the audiovisual sector.

The EIDR identifier and registry system is a new industry-led tool for digital distribution, whereas the ISAN is an established ISO Standard, primarily used for distribution of optical or magnetic media such as DVD and VHS. EIDR is still new, and is better known in the United States, though it is expected to grow in importance in Europe as it becomes established. ISAN is widely used, mainly in Europe but also in the US.

6.5.2 EIDR – Entertainment Identifier Registry

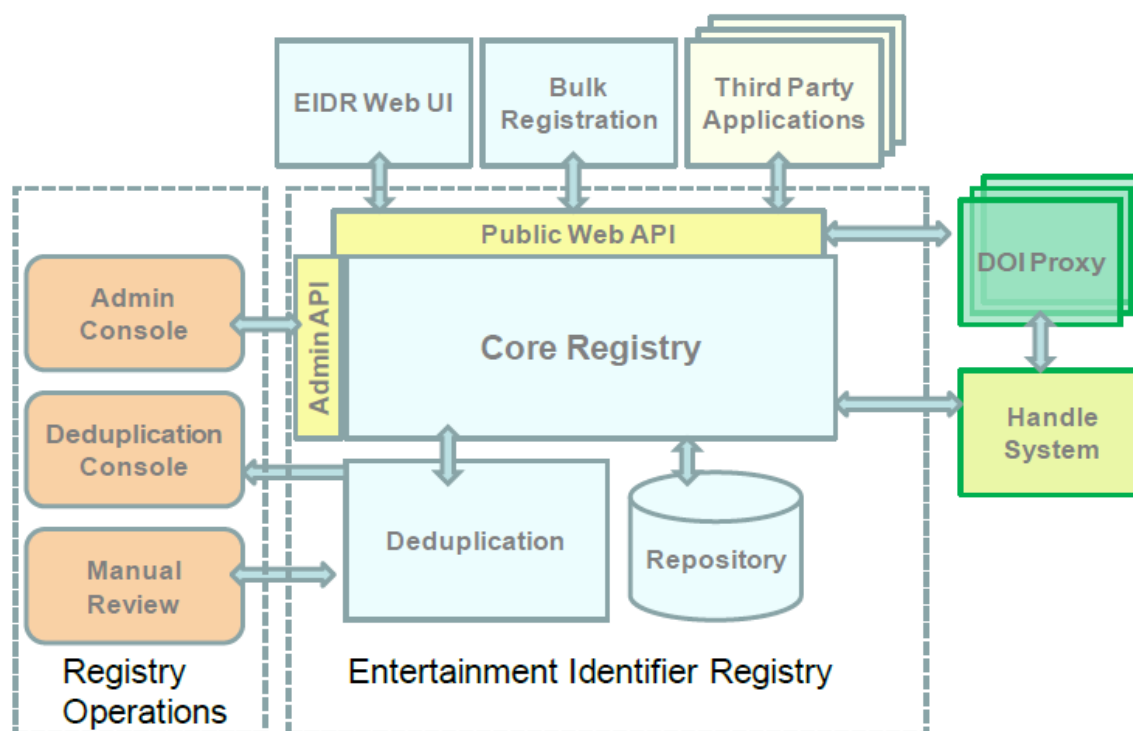
Introduction

As mentioned in the overview, EIDR (Entertainment Identifier Registry) has been developed deliberately by an industry body in the film business. It is an identifier, metadata framework and set of services based on the DOI standard and interoperable with existing standards such as ISAN (see below) and ISRC (see music section).

The EIDR is aimed at business-to-business use for distribution, retail and rights tracking; however, its services and systems, outlined in the following diagram from the EIDR technical overview (EIDR, 2010), lend themselves to support of other services such as metadata intermediaries by providing a key to link together data from various sources.

¹¹⁹ For more on the “Nordic model” see <http://www.kopinor.no/en/copyright/extended-collective-license/documents/the-extended-collective-license-as-applied-in-the-nordic-countries>

¹²⁰ See <http://www.movelabs.com/AboutUs/index.html>



Overview of EIDR system architecture

Scope

The EIDR data model, based on the conceptual framework of DOI (see table below), can describe audiovisual industry assets at all levels of abstraction, from the work through expression (e.g. when a movie was edited) and manifestation, and also in digital or “physical” (i.e. magnetic or optical) media (see EIDR, 2010, p. 9.).

The entities and relators in the EIDR data model can apply equally to televisual and cinema content, and born-Web assets, and are granular enough to encompass a range of ancillary assets, part and composite entities; for example, an entire feature film, clips and trailers based upon it, and the extras included in a particular DVD version thereof. The reference descriptive metadata set also includes a set of relator attributes for showing associations and dependencies. See the reference descriptive metadata section below for a diagram of these attributes.

The following kinds of entities are in scope for EIDR. Because EIDR uses a hierarchical data model with several levels of abstraction and specification it is not possible linearly to list possible in-scope items. Instead the following two tables show the top two levels of the hierarchy.

Structural types are inherited in EIDR from the basic DOI classification¹²¹. They could in principle apply to works and manifestations in any of the four media worlds in this report. The example “uses” here, tying them to the AV domain, are from EIDR, 2010:

Structural Type	Use
Abstraction	This is for is for objects that have no [perceptible] reality, such as a series container or the most basic concept of the original work.

¹²¹ See “DOI Metadata Kernel” at http://www.doi.org/handbook_2000/metadata.html#4.3.1

Structural Type	Use
Performance	This is for items that are a particular manifestation or version of something, such as the Director's Cut of a film or the Welsh-language version of a TV show.
Digital	This is a particular digital manifestation of a work, such as an MPEG-2 encoding of a movie.
Physical	This is for physical version of an object. EIDR will support this for physical films and tapes in a future release

From the above definitions it is clear that these types correspond to three distinct steps in the Index Model of Making:

EIDR	Index
Abstraction	Abstraction
Performance	Expression
Digital Physical	Manifestation

The “referent types” in EIDR describe AV-domain specifications of the above four general types:

Referent Type	Use
Series	An abstraction that contains ordered or unordered individual items
Season	A second level of grouping below a Series
TV	Content that first appeared via broadcast
Movie	Content that first appeared in a theater (in the US) or a cinema (in most of the rest of the world.)
Short	Loosely defined, to cover miscellaneous content such as published outtakes, special segments, <i>etc.</i>
Web	Content that first appeared on the Web. This is different from content from elsewhere that has been made available on the web.
Interactive Material	Assets which are not strictly audio-visual. It covers DVD menus, interactive TV overlays, customized players, <i>etc.</i>
Composite	Any asset composed of multiple other assets that is not more precisely describable.

In addition to these “facets” of object type, there is a final distinction between “basic” and “derived” types – the latter being more complex articulations of the former. The EIDR API Overview (EIDR, 2011) describes the allowed combinations of these types more fully. The derived entities are of course

more numerous and will not be described here. The important point is that they, along with their relationships to other entities, are considered equally as important as the basic types, to enhance uniqueness checking and usability.

The following relators allow specification of common types of AV material released along with more “substantial” content, linked to the main release by an EIDR ID:

Relator	Meaning(s)
“Adjunct” content	Interactivity, Outtake, Making Of, Interview, Music Video, Deleted Scene, Behind the scenes, B-roll, Featurette, Selected Clips, Other
“Alternate” content	Descriptive Audio, Camera Angle, Parental Control, Censored, Commentary (Director), Commentary (Other), Sing Along, Trivia Track, Other
Promotional content	Broadcast Ad, Theatrical Trailer, Infomercial, EPK, DVD Trailer, Web, Mobile, UGC Site, Radio Spot, Other

Governance

According to the EIDR website¹²², “The EIDR coalition is organized as an industry non-profit governed by a Board of Directors comprising key stakeholders from content owners, video post-production houses, content distributors, retailers, entertainment services and technology provider companies. Development of the registry is guided by a Technical Advisory Board consisting of representatives from members of the coalition”. EIDR is represented on the DOI Foundation board.

Construction

An EIDR asset identifier is a DOI identifier, with an ID type prefix (identifying it as a DOI within the Handle system¹²³), and then a unique suffix.

Element	Length	Possible values	Meaning
DOI identifier	2 digits	10	Denotes a DOI name
ID type prefix	4 digits	5237 5238 5240	Party (any entity with a role in creation, distribution or registration of assets) User (of the registry) Asset
Unique suffix	16 digits	0 – f	Dumb number
Check digit	1 digit	ISO 7064 Mod 37,36	Check character

In the format of this report the EIDR looks like this:

10.5237/hhhh-hhhh-hhhh-hhhh

10.5238/hhhh-hhhh-hhhh-hhhh-d

10.5240/hhhh-hhhh-hhhh-hhhh-hhhh-d

¹²² <http://eidr.org/membership/>

¹²³ For the Handle system, see <http://www.handle.net/factsheet.html> - full details are out of scope for this report

Reference descriptive metadata

The data elements presented here represent the minimum data possible for any object in the EIDR register; more complex “objects” are built from the “base object” specified here, and these will have their own mandatory and optional elements. The allowed values, as in the EIDR documentation, as expressed in XSD datatypes, which are clearly defined in the XML Schema definition¹²⁴.

Data elements	Cardinality	Allowed values	Comments
Structural Type	1	Abstraction, Performance, Digital, Physical	doi:creationStructuralType
Mode	1	Visual, audio-visual, audio	
Referent Type	1	Series, Season, TV, Movie, Short, Web, Composite, Interactive Material	doi:creationType
Resource Name	1	{xs:string of 0-128 characters, lang of xs:language, optional titleClass attribute enumerated}	The titleClass describes the title, which for the main title is usually “release”.
doi:creationName (with a Type of Title)			See explanation of types above.
Replaced Alternate Resource Names	1	xs:Boolean	For base objects this should be false.
Primary Language	1	xs:language	The main language used in the asset [e.g. the dialogue used]. For example, for the original releases of the Star Trek movies it would be English, not Romulan or Klingon.
IETF RFC 4646			
Principal Agent (or Producing Agent)	1-8	producer, distributor, broadcaster, encoding, post-production, editor	Used to identify the studio, producer, or other principal entity responsible for creation of the specific object

¹²⁴ XML Schema Part 2: Datatypes: <http://www.w3.org/TR/xmlschema-2/>

Data elements	Cardinality	Allowed values	Comments
Release Date	1	xs:gYear or xs:date	This can be specified as a year or as year, month, and day. When only a year is known, the month and day should be left blank. The term released may have different meanings in different contexts. For example, it is the actual release date for new movies; the original air date for TV shows; and the date a file was finished for a simple encoding.
Status	1	Valid or In-Development	“in development” records are hidden to everyone except the registrant or authorized proxies
Approximate Length	1	xs:duration	The approximate length of the object. For a Series and Season this should be zero (“PT0S”). The ApproximateLength field should be 0 for Interactive Material.
Administrators	1-n		Lists the registrant, and optionally either or both of the entities believed to be the current asset holder and a backup contact

This set of metadata is an overview – for the full set, the full EIDR schema should be consulted¹²⁵, and the data fields reference document for more details¹²⁶. Some examples of the structure of typical EIDR resources are outlined in the diagrams overleaf, again from the EIDR technical overview (EIDR, 2010). The MovieLabs *Common Metadata Schema* is also used in EIDR (see Appendix 9).

These diagrams highlight the available relators between entities, and also display the DOI types correlated against the relevant referent types to show how these are used together.

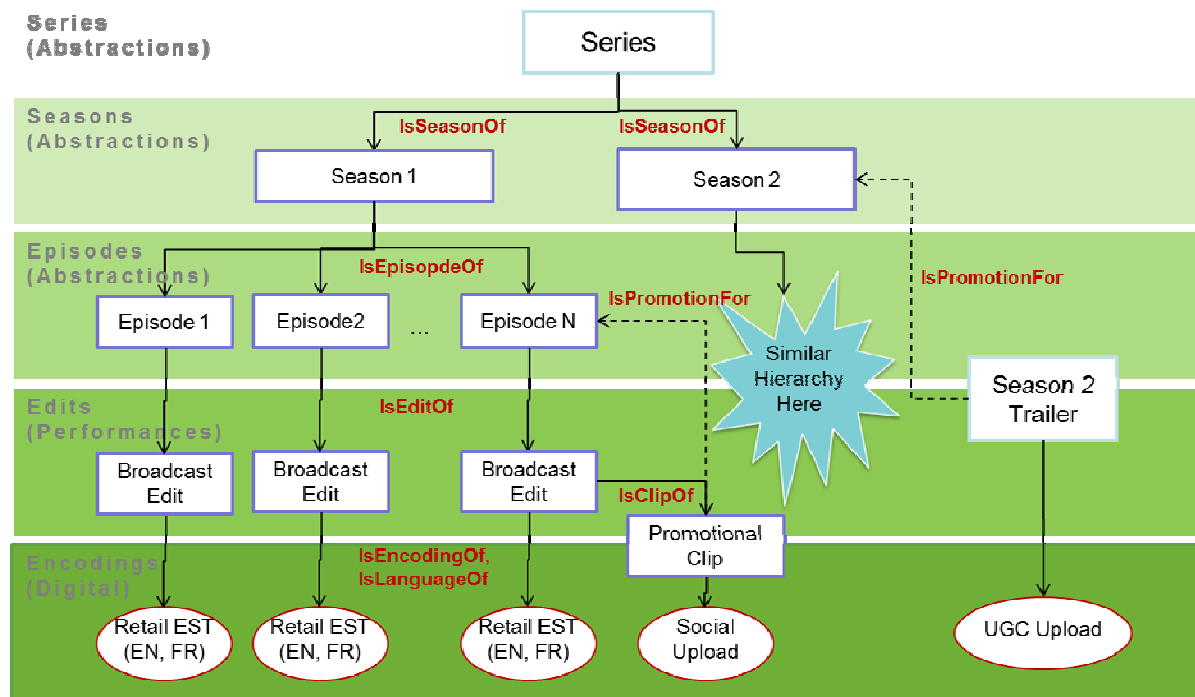
EIDR – Example resources and relations

The below examples demonstrate the possible relationships between entities at the three levels of abstraction identified above, e.g. the “encoding” attribute can only be used at the “digital” level.

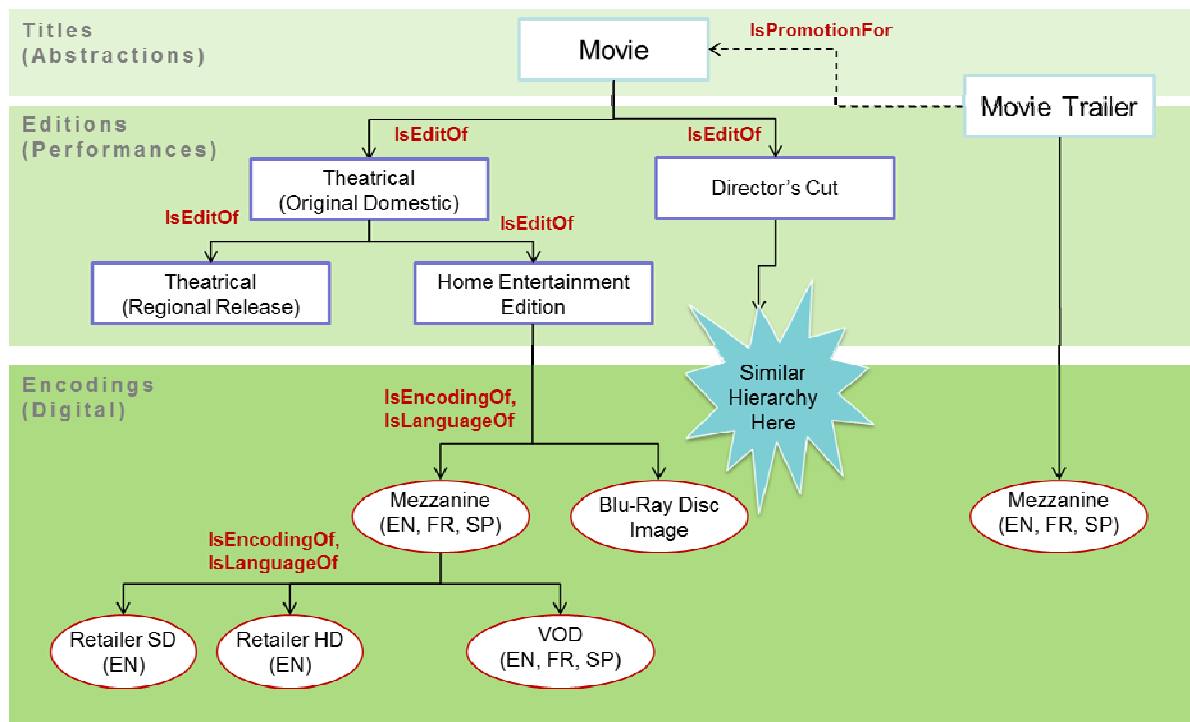
¹²⁵ See <http://www.eidr.org/schema/1.0/eidr-base.xsd>

¹²⁶ See http://eidr.org/documents/EIDR-1.03_Data_Fields.pdf

Example EIDR Episodic Hierarchy



Example EIDR Movie Hierarchy



The hierarchies above also demonstrate how the vitally important language attribute is inherited at the encoding level to allow a variety of distribution and packaging models, e.g. with different audio tracks and subtitles. Complex encodings of movie resources are also found in the technical overview (EIDR, 2010).

6.5.3 ISAN – International Standard Audiovisual Number

Introduction

The ISAN (International Standard Audiovisual Number) is a numbering system for audiovisual works. It is governed by ISAN-IA (the ISAN International Agency) in a similar way to the ISBN, ISMN and ISRC standards, in that ISAN-IA only allocates blocks of ISANs and responsibility for registration of AV works is devolved down to registration agencies at national level. An extended identifier, the V-ISAN, which specifies a version of a work (essentially an encoding and therefore a product identifier) has been described and put to ISO for approval.

ISAN-IA stated in 2010¹²⁷ that its database contained 600,000 registered numbers across 3500 registrants and 50 countries. This corresponded to approximately 70% televisual, 13% documentaries and 12% feature films, in direct contrast to the EIDR. The geographic range mainly covers North America (59%) with only 33% in Western Europe, and 4% from Eastern Europe.

Scope

The ISAN User Guide¹²⁸ lists the following types of work as eligible for ISAN registration:

- Commercials (including trailers, advertising films);
- Multimedia;
- Feature films;
- News;
- TV movies;
- Recordings of live (but planned) events, including;
 - o Documentaries;
 - o Performances (including music, theatre or other spectacles);
- Educational;
- Sports events;
- Serials (including dramas, comedies, soap operas);
- Other television entertainment;
- Shorts;
- Music videos.

The allowed “types” of work specified in the core (minimum reference) metadata are enumerated in the code list below (taken from the ISAN User Guide; International ISAN Agency, 2005).

Code	Type
AD	Commercials/Advertising film/Trailers
FF	Feature film
TF	TV movie or telefilm
DO	Documentary

¹²⁷ See ISAN newsletter, 2010: http://www.isan.org/docs/newsletter_1Q_2010.pdf

¹²⁸ See http://www.isan.org/docs/ISAN_User_Guide_2.2.pdf

Code	Type
ED	Educational
SE	Dramatic and comedy series, Serials, “Soaps”
SH	Short
LV	Live event
PF	Performance
SP	Sporting event
TE	Other television entertainment
VC	Music video clip
MM	Multimedia
NE	News

Construction

The ISAN uses hexadecimal digits. It consists of a 16-digit root identifying the work, of which the final four digits optionally identify the episode being identified (if not they are all set to 0 by default). A checksum digit follows.

Element	Length	Possible values	Meaning
Root	12 digits	0–f	Work identifier code
Episode	4 digits	0–f (0000 if not serial work)	Episode identifier code
Check digit	1 digit	0–f	Work check digit
Version number	8 digits	0–f	Version identifier code
Check digit	1 digit	0–f	Version check digit

As shown below by the section in [square brackets], the extended version identifier, the V-ISAN, can be constructed by adding 2 further blocks of four hexadecimal digits each, followed by a version code checksum digit. This V-ISAN is, however, pending ISO approval and so is not yet in use.

The episode section is shown in CAPITALS.

ISAN hhhh-hhhh-hhhh-HHHH-c-[hhhh-hhhh-c]

Reference descriptive metadata

The groupings here are based on the presentation of the elements found in the ISAN Handbook.

Block	Data elements	Cardinality	Allowed values	Comments
Core data	Type	1	Code list – see scope above	
	Kind	1	Live action, Animated, Combination of both	
	Year of reference	1		
	Duration	1		
	Composite	1 if applicable	Title ISAN	“Supply the title and ISAN of each component part, if applicable and available”
Production information	Coproduction	1	Yes or no	
	Original language	1-n	Code list, none	
Titles	Original	1	Yes or no	One selection for each title
	Language	1	ISO 639-2	One language code for each title
	Title	1-n		These can be episode titles within a series
	Alternative titles in the original language if applicable	0-n		mandatory if applicable
Participants	Type (Director)	1		
	Type (Actor)	3-n if applicable		“At least three, including First and Last names... if applicable and if available... lead characters’ names may be used if needed”

Block	Data elements	Cardinality	Allowed values	Comments
Episodes	Is Episode From	1	Title of the parent work	This and the field below appear for episodes; the other fields above for “work” are also repeated.
	Episode Number	1		

6.5.4 Technical standards and DRM: MPEG-7 and MPEG-21

MPEG

The Motion Picture Experts Group (MPEG)¹²⁹ sets technical standards for the audiovisual content industries, mainly focusing on encoding / decoding of digital content, but also including descriptive metadata standards such as MPEG-7¹³⁰, and the MPEG-21 “open framework”, which is why it merits mention here. MPEG standards are ISO-led, but also relate to film industry-led DRM standardisation attempts described below.

Anti-piracy and DRM

Motion picture and TV piracy was already a concern for these industries in the early stages of personal video camera and recorder development. In this phase of digital content and the Internet, cultural industries are looking to technical solutions like DRM (digital rights management) to enable distribution that is both convenient and attractive to wide markets and also protected from copyright and license infringement. Some of these initiatives include “rights lockers” like Ultraviolet¹³¹ and Keychest¹³² which, being secure registries of customer identifiers and access rights, necessarily involve some standard metadata.

6.5.5 Descriptive and marketing metadata

While there is no single source of current filmographic and television data similar to the concept of the books-in-print database (see section 6.3.1.) in the book publishing world, several commercial databases approach to this and are worthy of mention here for their coverage, although they do not aim to present standards in metadata format nor identification methods.

IMDB

The Internet Movie Database (IMDB) is a commercial service now owned by Amazon, although it began as a free, user-created database of movie information¹³³. It includes TV listings information and so overlaps in coverage with Rovi.

TV Guide / Rovi

¹²⁹ See the MPEG homepage for more details: <http://mpeg.chiariglione.org/>

¹³⁰ For an overview of MPEG-7 see <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm> and for the MPEG-7 schema see http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-7_schema_files/mpeg7-v2.xsd

¹³¹ See <http://www.uvu.com/>

¹³² See <http://online.wsj.com/article/SB10001424052748703816204574485650026945222.html>

¹³³ See <http://www.prnewswire.co.uk/cgi/news/release?id=37602>

Rovi Corporation delivers metadata for music, TV and film¹³⁴, in both consumer and business-to-business service modes. Other metadata providers similar to ROVI include IVA (Internet Video Archive) and Baseline; Red Bee Media and Tribune Media are important providers in Europe.

Findanyfilm

One example of a current film metadata aggregator with links to buy products is Findanyfilm.com, based in the UK and founded by the UK Film Council (now part of the British Film Institute¹³⁵). There is a minimal set of descriptive metadata, trailer (preview) and availability by cinema, retailer and online distribution.

6.6 Photography

6.6.1 The photography sector

This sector contrasts profoundly with the other three in that its products, unlike books, films and music recordings, are rarely sold directly to the public. They are normally sold to other commercial companies within another supply chain such as news or education publishing, and incorporated into other creative works – for example, newspapers, magazines, news or entertainment websites, advertising or textbooks – before reaching the retail stage.

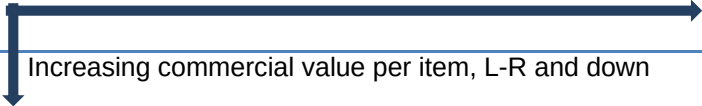
This contrast is clearly seen in the fact that ISO standard identifiers exist for all three others, but not for images; and the fact that most metadata for commercial images are “in-band”, that is, incorporated into the electronic media file itself throughout the supply chain, rather than transmitted separately from the pictures.

It is possible to distinguish somewhat distinct markets for still images defined by their level of specificity (“stock”, or generic, versus directly commissioned) and the secondary commercial use to which they put (“editorial” or news, versus advertising or “product placement”).

Each market also lends itself to specialization – news picture agencies need to be focused on current affairs and thus extremely *specific and unique images*; in contrast, images for book covers (and to some extent, illustrations), advertising and informational brochures *etc.* need to be relatively *generic and unidentifiable* especially where they contain human models or commercially sensitive designs or patented technology. Hence there are companies specialising in *news photos* and those that tend towards *stock*. This analysis is summarised below:

¹³⁴ See http://www.rovicorp.com/products/online_stores_portals/11242.htm?link_id=rightnav

¹³⁵ See <http://www.bfi.org.uk/about/policy/bfi-ukfc-transfer.html>

	Secondary commercial use		
Specificity of requisition	<i>Book publishing (covers, illustrations etc.)</i>	<i>Editorial / news</i>	<i>Advertisement / product placement</i>
Stock			
Commissioned			

Of course there is some overlap, as mentioned above, and each of these categories could contain an online element, with “new media” markets such as “monetized”, commercial blogging, in particular, crossing the boundaries between news press and advertising. “Product placement” refers to the use of an advertising-oriented image in (broadly speaking) an editorial or informational context; generally this occurs in consumer magazines and retail catalogues. Generally, more specific photographs will cost more, and the greater the expected return on the final commercial product, the greater income can be expected from the sale of the image.

A final category, *historic* images, might be suggested for *archival images* held by picture libraries and archives, and licensed for use in books, design, advertising and the cultural heritage sector. This is mostly *commercial* but *not for profit*. An even more specific and unique field is that of *art photography*, high priced¹³⁶ and extremely rare¹³⁷ photographic objects considered as fine art. This final category is out of scope for this report.

On the other hand, there is a very strong tendency for horizontal media sector integration among larger organisations, with companies that were originally photo agencies acquiring or creating libraries of “stock” parts of other media, especially online:

- Getty Images (Gettyimages.com) – also sells “stock” music and (AV) footage;
- iStockphoto (iStockphoto.com) – also includes illustrations, video, audio and Flash animations;
- Corbis (Corbis.com) – also features illustrations; other Corbis sites include:
 - o Corbismotion.com – AV;
 - o GreenLightRights.com – music, film and TV rights clearances.

Diversification of offering in these companies could be seen as an example of a common media industry strategy (Albarran, 2010) or potentially a response to the threat of competition from high-quality amateur photography enabled by the Internet and Web 2.0 (Johnson, 2011). As explained by Johnson, a new sub-sector has evolved in the Internet age, that of “micro-stock”, with significantly lower costs to users and proportionally lower rewards for contributors, and a multiplicity of rights schemes and remuneration systems.

6.6.2 Outline workflows in photography industries

In this industry sector, as noted above, data are added directly to file format wrappers from the moment of creation of the photograph (for example, the data and time, automatically added by digital

¹³⁶ See FT.com’s analysis of recent art photo prices:

<http://arts-extra.ft.com/artprice/arts-price-index-23-photographs-2004-2011>

¹³⁷ See, for example, <http://www.ft.com/cms/s/2/54b5c4f6-f81a-11df-8875-00144feab49a.html> on art photo collecting.

cameras). More detailed data are added at every stage of the photo's lifecycle, depending on the type of photography.

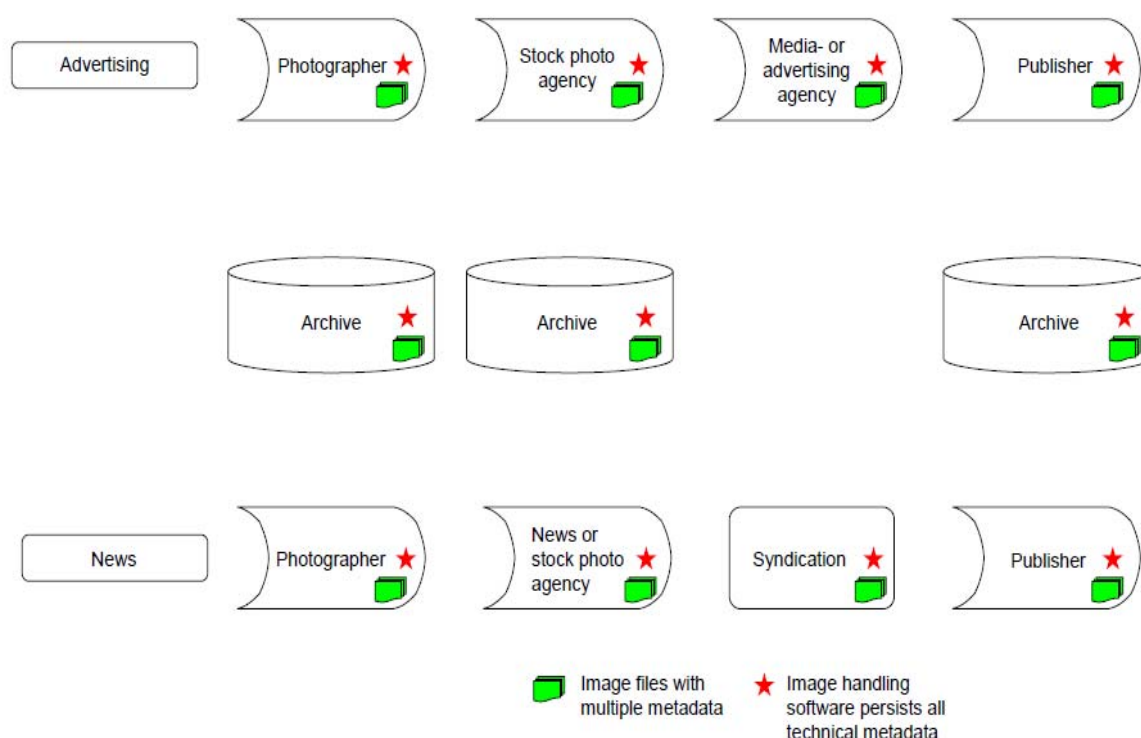
In-band versus extraneous metadata

As mentioned above, the data are almost all “in-band” which makes the dynamic nature of commercial metadata especially apparent in this industry. The data held by academic or other research institutions, cultural heritage organisations exists extraneously to the image files; a commercial analogy exists in picture libraries which sell their images, but these by the same token do not have such high turnover of files and metadata. They also tend to use in-house formats and release data only via API (e.g. to partners who host search boxes or related search results). This approach reflects the lack of a mass retail market beyond very relevant niche partners.

A major concern for most companies, therefore, is the possible deletion of essential metadata at any point in the chain, often in practice at the point of incorporation into a composite work and publication; and the loss of potential secondary earnings from re-use of royalty (pay-per-use) or relicensed images.

Generalised photography business supply chains

The “generalised photo value chain” below (from IPTC, 2010, p. 52) shows, moving from left to right, the flow of photographs in 2 distinct industry sub-sectors; advertising and news. Between the two are indicated the three points where archives may be maintained of items from either industry, suggesting there can be interaction between these two types of media work via the archives (it is not unknown for news media to carry advertising, or to themselves be advertised in other publications, for instance).



The purpose of this generalisation is to highlight the need to “persist” metadata along the entire flow and the points where three main reasons for metadata loss may occur¹³⁸:

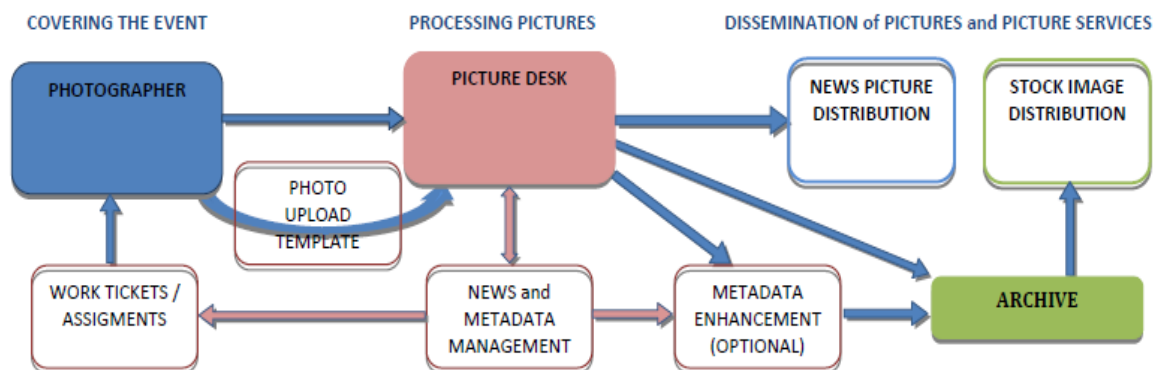
- Human intervention;

¹³⁸ See MWG, 2010 for detailed discussion of how metadata can be lost by human or automated “agents”.

- b) Software settings;
- c) Change of image file format (e.g. TIFF to JPEG).

News photography example workflow

The diagram below outlines how metadata is created by the picture desk, chronologically even before pictures are taken, as it must be aligned with other news business processes (CEPIC, 2011).



The table below summarises where CEPIC identifies addition of data and some extra details in [square brackets]:

Supply chain stage	Data added to picture
Picture desk	Job number News / contextual data
Photographer	Photo content data [GPS information] [camera technical metadata] [photographer identification]
Archive	Keywords Supplier name PLUS licensing data

Stock agencies example supply chain

The situation is drastically simplified, as shown below (again from the CEPIC, 2011, handbook) resulting in a different set of metadata:



Supply chain stage	Data added to picture
Photographer	Creator Headline Description Some keywords Terms of use
Stock agency	Keywords Credit line Rights data

It will be noticed that rights information does appear in the processes above. However, as licences are not machine-readable the metadata is not as valuable as it could be in business-to-business transactions. Hence there is currently interest in developing rights expression languages such as RightsML¹³⁹ and ODRL¹⁴⁰, though none is yet adopted as a standard.

6.6.3 Rights and licensing in and of photos

Commercial photos are associated with two particular legal issues worthy of note here:

1. They mainly contain locations, branded objects (and c.f. “product placement”) or human models who appear in the photograph under contract that specifies the terms of use of the picture – and conversely, the pictures themselves are very often (re-)used in other creative works; thus they can be part of complex configurations of works and rights up and down the supply chain;
2. Rights and licensing information is not easily incorporated into the image in a human-readable form as it might be on a book’s title page, a sound recording as speech or text on the cover, or as a still frame in an AV recording. There is therefore a higher intrinsic risk of metadata loss and creation of new orphan works.

Historically there is no commercial supply chain identifier in use for either analogue or digital photos, leading to a clear risk of orphan works. A specific picture identifier could potentially be linked to identifiers for composite works such as texts (ISTC) and AV (EIDR or ISAN) or musical items (e.g. GRid) and allow tracking of rights (for remuneration) and description (for discovery) through multiple uses.

6.6.4 Technical creation and reproduction of photo metadata

In the photography sector several layers of more-or-less standardised metadata are added to digital photographic files. It is useful to understand the broad outlines of this process as they affect what metadata may be present in commercial sector contributions to Europeana, and the way that this data may be processed.

In the digital camera industry, including related devices such as electronic picture storage and photo printers, the *de facto* technical standards are:

- DCF (Design rule for Camera Filesystem)¹⁴¹;

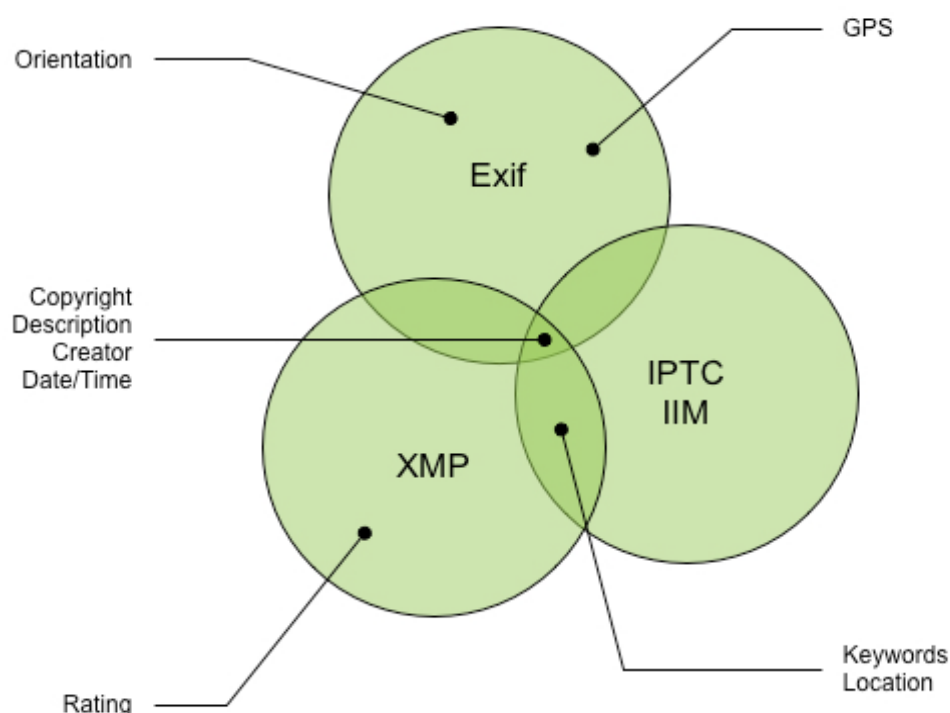
¹³⁹ <http://www.w3.org/community/odrl/files/2011/07/IPTC-20110928-RightsML-ODRLmeeting.pdf>

¹⁴⁰ <http://odrl.net/>

¹⁴¹ Both DCF and Exif are found at http://www.jeita.or.jp/english/standard/html/1_4.html

- o Specifies how image files and related thumbnail previews should be identified and stored;
- o Data in this format largely consists of the image file itself, but preview thumbnails could conceivably be considered “metadata” in the sense of collateral content described in previous sections;
- Exif (Exchangable Image File format);
 - o Specifies technical, descriptive and rights metadata related to the photographer, camera and image.

Here it should be noted that from the beginning of the supply chain, technical, administrative, descriptive and rights metadata are intermingled. The Exif and DCF metadata are normally combined together by photo storage and processing software, and this composite then partially overlaps with the content supply chain metadata standard, IPTC/IIM, and the Adobe-developed standard XMP, as illustrated by this diagram:



Image

source: MWG, 2010; p. 9.

The above complex of metadata types and schemas is further complicated in practical situations by the fact that the various proprietary software applications use slightly varying internal metadata formats and thus tend to either need mappings to interoperate or will mismatch or lose certain fields depending on the combination. One overview of popular software data models and elements is provided by the Stock Artists Alliance (SAA) Photo Metadata Project (PMP)¹⁴².

The concept of “standardising” metadata workflows within the individual photographer’s studio or field equipment environment is thus accepted as one way to cope with the multitude of formats and combinations of software approaches. In the absence of a fully-developed, central data repository such “standards” are of course highly personalised.

¹⁴² See the SAA element list here: <http://www.photometadata.org/META-Resources-Field-Guide-to-Metadata> and the crosswalk there referenced: http://www.controlledvocabulary.com/imagedatabases/iptc_core_mapped.pdf

6.6.5 Identifiers for digital photographs?

There is currently no unique identifier standard for photographs. This is clearly due in part to the nature of the photography industry supply chain, embedded within other supply chains where the photograph loses its identity as a distinct work. Another factor may be the ease with which an in-band digital identifier might be lost due to deletion of metadata in the digital file at application interfaces, file format transformation, publication *etc.*

The lack of a whole-sector identifier means that trading of photographs through multiple supply stages is not efficient or reliable – of course, this is unproblematic in the traditional business-to-business supply chain. However, digital reproduction has made re-circulation of high-quality originals much easier, and digital publishing means that creation of derivative works including one or more rights-controlled photographs is much simpler. These factors combined make photograph professionals uneasy and create concern for the proliferation of orphan works and lost revenue.

- The IIM (see below) had provision for a UNO, or “Unique Name of Object”, which unfortunately was partially assigned by the object’s creator and not apparently registered with any independent body and so not persistent. The IIM specification is also no longer in development.
- The Adobe XMP specification described below recommends the use of GUIDs (Globally Unique Identifiers) which are a de facto standard promoted by the Internet Engineering Task Force¹⁴³ as a cost-effective alternative to identifier registration. They are effectively very large random numbers, so although not guaranteed unique, they are statistically (almost) impossible to duplicate so the chances of an identification mistake become vanishingly small. Although this type of identifier is clearly suited to processes involving high numbers of original works which may go through hundreds or thousands of iterations, it remains to be seen if they can gain broad professional acceptance or official recognition as a de jure standard.
- One step towards an industry-wide image identifier is the PLUS license standard¹⁴⁴ and registry¹⁴⁵, which include (optional) IDs for photo creators, agencies and licensors. This is intended to become an interoperable linking identifier similar to the EIDR in the AV sector, based on the CNRI Handle system (like the DOI). Its registry incorporates three safeguard mechanisms to ensure identification of parties, creators and pictures:
 - o Persistent identifiers and updated metadata;
 - o Digital watermark identifiers for images;
 - o Image search.

The PLUS framework also includes standardised terminology and semantics for picture license terms, media types and metadata elements for photo description.

The PLUS Registry is currently in a beta phase of development¹⁴⁶ and is so far only taking registrations for organisations. It will require significant take-up from major industry players in all of the main sectors, with their different emphases and business needs, to become plausible as an ID and metadata framework for large-scale aggregation of data. This could take a significant number of years, probably beyond the timescale of the Linked Heritage project.

6.6.6 IPTC and XMP

Overview of IPTC and XMP

This report treats IPTC and XMP metadata as equally important, since although the IPTC standard is a rigorous standard defined and endorsed by the professional photography community (in the form of

¹⁴³ Full definition and description at <http://www.ietf.org/rfc/rfc4122.txt>

¹⁴⁴ See <http://www.useplus.com/useplus/standards.asp> – free registration required to access

¹⁴⁵ See <https://www.plusregistry.org/cgi-bin/WebObjects/PlusDB>

¹⁴⁶ See the July 2011 article “On the PLUS side” <http://www.epuk.org/Opinion/985/on-the-plus-side>

IPTC and CEPIC), its practical software implementation in XMP has become the accepted means by which it is used. Because of the complexity of the composite XMP it is worthwhile to summarise the chronological processes (within each record and in terms of standards) that produce it.

A CEPIC/IPTC metadata survey carried out in 2010 “found that 70% of image suppliers were using IPTC fields as specified” (CEPIC, 2010). For the current task, a representative sample of commercial picture libraries and agencies, previously examined in the JISC (2006) CLIC project, were polled directly to ask if they used IPTC. All but 2 respondents (see Appendix 6) confirmed they did.

Exif

Exif is the fundamental technical metadata added by the digital camera software on creation of the picture. It is only noted here because the Exif data will be present in most digital image files – although an important caveat is that some software deliberately or accidentally deletes portions of Exif. The full Exif documentation is available online at the website of JEITA (2010) and commentaries exist¹⁴⁷.

IIM and IPTC

Historically, the IPTC standard is based on an earlier group of metadata elements, the Information Interchange Model (or IIM), a multimedia standard defined by the IPTC and the NAA (Newspaper Association of America).

Adobe XMP

The diagram below illustrates at the top level how Exif and IPTC are embedded within the XMP associated with a JPEG image. It also shows:

1. Additional XMP elements;
2. Descriptive Dublin Core elements;
3. Legacy IIM elements (see section on IPTC below);
4. Custom, optional elements added within a software package by a user.

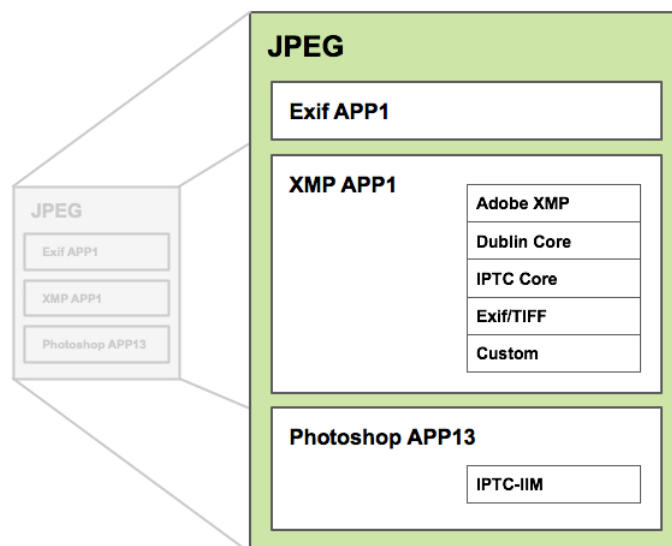


Image source: MWG, 2010; p. 9.

The tables on the next two pages expand the overview of XMP further, first showing the mapping of IPTC Core onto XMP along with the photo industry terms for each element, then the IPTC Extension

¹⁴⁷ See

<http://web.archive.org/web/20110716154113/http://park2.wakwak.com/~tsuruzoh/Computer/Digicams/exif-e.html> for explanation of Exif

fields (including PLUS rights and licensing elements) mapped onto XMP. The IPTC extension should also be shown in the MWG's table above under the XMP APP1 block.

This table highlights the way that IPTC Core and Extension and other vocabularies' elements are incorporated into the XMP wrapper structure. The short note on each data element's semantics has been provided by our industry contributor. For a full set of official definitions see IPTC, 2010. There are also more detailed guidelines on each of the IPTC Core and Extension fields available from <http://www.iptc.org/std/photometadata/documentation/GenericGuidelines/>.

IPTC Core Generic Name	XMP namespace:property ID	Definition (the semantics of this field – how to use/what to put in here)
Creator	dc:creator	Byline
Description	dc:description	Caption
Copyright Notice	dc:rights	Copyright
Keywords	dc:subject	Keywords
Title	dc:title	Normally used to store the filename of the image; not the same as Headline.
Country Code (legacy)	Iptc4xmpCore:CountryCode	Country ISO Code (ISO 3166-1 alpha-3)
Sublocation (legacy)	Iptc4xmpCore:Location	Specific location, such as Madison Square Garden, Kodak Theatre, Grand Canyon etc...
Creator's jobtitle	photoshop:AuthorsPosition	Byline Title
Description Writer	photoshop:CaptionWriter	Initials of Caption Writer
	photoshop:Category	Category
City (legacy)	photoshop:City	City
Country (legacy)	photoshop:Country	Country Name
Credit Line	photoshop:Credit	Credit
Date Created	photoshop:DateCreated	Create Date

IPTC Core Generic Name	XMP namespace:property ID	Definition (the semantics of this field – how to use/what to put in here)
Headline	photoshop:Headline	Headline / Event Name
Instructions	photoshop:Instructions	Special Instructions
Source	photoshop:Source	Source
Province or State (legacy)	photoshop:State	State or Province
	photoshop:SupplementalCategories	Supplemental Categories
Job Id	photoshop:TransmissionReference	Media Event ID (7-9 numeric chrs)

IPTC Extension	XMP namespace:property ID
Additional Model Information	lptc4xmpExt:AddlModelInfo
Artwork or Object in the Image	lptc4xmpExt:ArtworkOrObject
Code of Organisation Featured in the Image	lptc4xmpExt:OrganisationInImageCode
Controlled Vocabulary Term	lptc4xmpExt:CVterm
Location Shown in the Image	lptc4xmpExt:LocationShown
Model Age	lptc4xmpExt:ModelAge
Name of Organisation Featured in the Image	lptc4xmpExt:OrganisationInImageName
Person Shown in the Image	lptc4xmpExt:PersonInImage
Digital Image GUID	lptc4xmpExt:DigImageGUID
Digital Source Type	lptc4xmpExt:DigitalSourceType
Event	lptc4xmpExt:Event
Image Registry Entry	lptc4xmpExt:RegistryId
Image Supplier	lptc4xmpExt:ImageSupplier

IPTC Extension	XMP namespace:property ID
Location created	lptc4xmpExt:LocationCreated
Max Avail Height	lptc4xmpExt:MaxAvailHeight
Max Avail Width	lptc4xmpExt:MaxAvailWidth
PLUS Version	plus:Version
Copyright Owner	plus:CopyrightOwner
Image Creator	plus:ImageCreator
Licenser	plus:Licenser
Minor Model Age Disclosure	plus:MinorModelAgeDisclosure
Model Release Id	plus:ModelReleaseID
Model Release Status	plus:ModelReleaseStatus
Property Release Id	plus:PropertyReleaseID
Property Release Status	plus:PropertyReleaseStatus

Based on an original contribution (see Appendix 4) by Andreas Gebhard.

It should be noted that the XMP is a composite of Dublin Core, IPTC and PLUS elements, together with some elements from the Adobe image manipulation software application PhotoShop. This means that it is already reliant on other namespaces, some of which have overlapping or ambiguously defined meanings.

The IPTC (International Press Communications Council) created the IPTC Core and Extension metadata schemas to standardise communications within the press photography supply chain (they existed alongside a more generalised, legacy model – the IIM mentioned below – for all news content including text, pictures and audio). Because the picture creators and agencies needed to create efficient communications horizontally the standard is useful to those outside the immediate news press workflows; however, it is not based on a coherent conceptual model adapted to a wide variety of use cases. Sample IPTC/Exif/XMP records can be found in Appendix 5.

Another widely used framework, PRISM (Publishing Requirements for Industry Standard Metadata) has a distinct image specification, PRISM Metadata for Images (PMI)¹⁴⁸. This shares some fields with IPTC Core and Dublin Core, but as it is intended for use solely within the news and magazine publishing supply chain it is out of scope for this project, and being embedded within publishing work objects, has no status as a supply standard.

Governance of IPTC and XMP

The IPTC Core and Extension metadata elements are maintained and governed by a committee of the International Press and Telecommunications Council.

¹⁴⁸ Full documentation on PRISM specifications including PMI at <http://www.idealliance.org/specifications/prism/resources/46>

XMP is a proprietary Adobe standard entirely governed by that company, although the specifications are openly available.

7 Legal-commercial implications of metadata re-use

Although this deliverable does not explore the commercial-legal aspect of commercial sector metadata (that will be covered in Task T4.2 and Deliverable D4.3) it is unavoidable to mention some of the basic facts of the current situation, as these are an integral part of the landscape of commercial sector metadata.

7.1 Drive for openness of public sector metadata

There is currently a widespread tendency for public bodies to open up their data to public access. The motivations and justifications for this include:

- Public access to valuable data whose collection and compilation is publicly-funded;
- Transparency and accountability of the government and public services;
- Cultural interests of citizens;
- Accuracy of data through visibility and mechanisms for updating or correcting;
- Commercial interests of companies who can re-use and build services on data (but see below!);
- Benefits to public image of originators of data due to increased awareness of their output;
- Enrichment of data for all of the above purposes through linking datasets.

Similar reasons have inspired the decision to open up all of Europeana's object metadata to the public under a CC0 license and publish it as "open linked data".

This report, aiming to represent and interpret the commercial sector's interests to our (mainly public heritage sector) partners in Linked Heritage, colleagues in Europeana and beyond, appreciates and accepts the reasoning behind opening up publicly-funded, public sector metadata.

However, for the same reason, it must also recognise the other side to this story – commercial sector metadata is not produced at public expense; it is potentially commercially sensitive; it may be protected by database right, and it may be a significant commercial asset as a corpus of structured, managed data. Linked data is not, and does not always have to be, (entirely) open data.

7.2 Legal status of commercial metadata

Metadata itself, even in the commercial sector, is probably not protected by copyright, since it is basically descriptive of facts and in principle easily created by someone else¹⁴⁹. Previews and extracts, as mentioned above, go beyond this concept of (textual descriptive) metadata and are not at issue here.

7.2.1 Metadata normally make up databases

Metadata *per se* tend to be applied to large quantities of data and so make up substantial and normally highly structured databases. European law, implemented through Directive, does in fact protect databases (e.g. Dierickx, B. and Vissers, R., 2009):

"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."

¹⁴⁹ See e.g. legal opinion based German law available at <http://www.version1.europeana.eu/web/europeana-project/newagreement>

7.2.2 Databases are commercially significant

Raymond Bérard emphasises this point in his article *Free Library Data?* (Bérard, 2011) and notes also:

- 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 ongoing 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*.

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 6.3.1.) and media listings (see section 6.5.5), are already bought and sold. It also has a role in making work processes more efficient, and thus has an internal economic value to companies.

7.2.3 Confidentiality and implied licenses

There may be a perceived or real competitive advantage in keeping metadata secret from other companies in the same sector, or at least in retaining control of when and how it is released. Although increasingly metadata is seen as key to marketing and strategic decision-making, existing identifier and metadata frameworks are normally based on *closed data* models – where some parts of the data may be publicly shared, much remains internal to companies or third party registries, as illustrated in the example of ISNI's structure below:

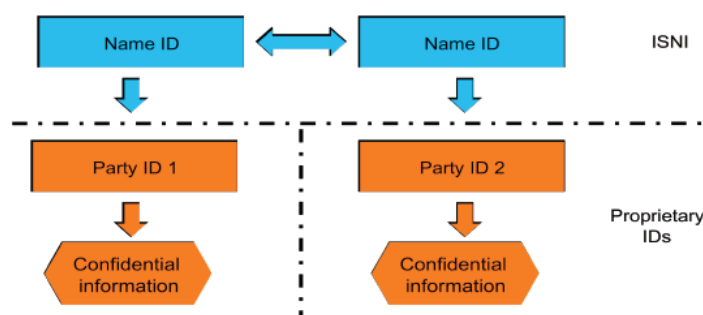


Diagram from International ISNI Agency, (2009).

Although the ISNI case of confidential, personal data is an extreme one, it illustrates concerns that are common in the commercial sector; entities can be collocated without merging the separate bodies of information. An explicit agreement when substantial metadata is traded between commercial companies commonly restricts its use to within the company buying the data, keeping it from hidden consumers of the retail product.

Even when an explicit licence for use of the metadata is absent, the concept of an *implied license* is also relevant here (c.f. Bell, 2011 – relevant extract supplied in Appendix 7), as in the terms of use created by the intentions for which metadata has been supplied to an organisation (e.g. an aggregator) as reasonably inferred from the obvious commercial aims of the supplier, the normal operations of the aggregator, and the nature of the data. This *intention* will only extend to use for marketing and promotion, not redistribution, commercial or otherwise.

7.2.4 Territorial sales rights and windowing

The timing of the release of metadata is important in the commercial world, as it relates to the marketing function of the data. Descriptive data may be updated right up until the day of release. In some sectors, especially film and TV and music, but increasingly others as more content is distributed digitally, products are released in particular formats and to different markets in a strict chronological sequence (“windowing”), and metadata releases may be synchronised (“day and date”) with other releases of products and metadata.

The different markets may be defined geographically or territorially, and the right to sell the product in that territory or market is controlled – therefore the content and timing of data released about the product will also be tailored to each market in turn. Making one generic message about all formats and territorial variations of a product will not be acceptable in such cases.

The intention here is not to provide a complete overview of commercial metadata in legal and business terms, but to situate the findings here in their proper context of commercial use and existing legal framework, and to provide indications for future work. The current situation is based on current assumptions and agreements; new partnerships and collaborations will require their own negotiations and will be based on new terms.

8 Existing partnerships

Linked Heritage aims to build on approaches already existing in the European Community, and to make recommendations for partnerships in harmony with existing collaborations (or at least learn from the best attempts made to date). Several clear indications are available in each of these areas.

8.1 Partnership Approach Assumed in Linked Heritage

The FEP approach described in section 2.3.1., informs the structure of this entire report and the envisaged next steps. Since no content can in any case be hosted in the Europeana system, and commercial providers would not offer their products for free, some way of linking the metadata that will appear in Europeana with a persistent link to an online retail portal must be found – this is precisely the sort of use case that most of the identifiers and metadata schemas presented here address.

- Rich metadata – is key to ensuring discoverability of commercial items, demonstrating their interest to Europeana users and making it clear that these are paid-for materials;
- Robust identifiers – are essential to ensure that the correct version of a product is found and requested, and that links to the retail level resolve correctly.

8.2 Proofs of concept

There are existing public-private partnerships in two of the areas this report addresses:

8.2.1 Books: Gallica 2 (France)

The National Library of France in 1997 created a digital library which contributes metadata to Europeana. Since 2006 it has run a program to assess the business case and legal framework for including in-copyright books' metadata and links to buy the full item. As well as metadata and retail site link, Gallica 2 also enables full-text upload (for search only, not display), as is the case for Enclave, detailed below. Their consultant (Numilog) addressed the following points¹⁵⁰:

1. Reliable protection of documents in copyright;
2. Appropriate payment for copyright holders;
3. Integration of the different players of book industry;
4. Simple and easy access for the Web user;
5. No exclusivity: material in copyright can also be displayed by other portals (publisher's choice);
6. No commercial role for BnF (including exclusion of affiliate payments for retail links).

8.2.2 Books: Enclave (Spain)

This collaboration of the National Library of Spain with the Spanish Publisher's Association, to use DILVE, the Spanish books-in-print repository, as a source for in-copyright book metadata for the online catalogue of the National Library, and include links to the publishers' website for purchase of the full book. The data provided was the following ONIX for Books elements for each product¹⁵¹:

- Main description (plain text format);
- TOC (plain text format);
- Cover image (JPG or PNG);

¹⁵⁰ All details in this section from Girard, 2008.

¹⁵¹ All details on Enclave from Peraita, 2010.

- Excerpt for on-site browsing (PDF);
- Full text PDF (Internal BNE use; for indexing only);

and also (not provided via ONIX, although this would have been possible):

- One URL to an e-distributor / e-retailer environment.

In contrast to Gallica 2, which is running as a stable service, Enclave ended as a project in 2010. It produced the following learning points:

- Full text indexing via PDF file (BNE requirement) problematic for books in EPUB format;
- URLs for e-distributors are very volatile.

8.2.3 Film & TV: Ximon (Netherlands)

The Ximon joint initiative of the Dutch Film Institute and Dutch Feature Film Association combines archive film and TV (free to view) with current commercial Dutch films¹⁵², which through “agreements with copyright collecting agencies... for this model of distributing video content”, are discoverable through metadata and previews (trailers) and purchased one at a time or by subscription, via direct online streaming.

Notable features of Ximon’s business model are:

- Commercial videos (and even their trailers) cannot be streamed outside of the Netherlands;
- Embedded videos usable on other sites are in fact linked back to Ximon’s site.

8.2.4 Film & TV: FindAnyFilm.com

The Find Any Film website is another partnership in the audiovisual sector. It was originally founded as a promotional initiative by the UK Film Council¹⁵³, whose responsibilities have now been transferred to the British Film Institute (BFI)¹⁵⁴.

- aggregates filmographic information from a variety of sources;
- integrates sources into records with an internal identifier;
- pulls in trailers and advertising from third-party services;
- for each movie, displays;
 - o cinematic release details if current (with location search);
 - o links to retailers to buy movie in various formats (e.g. Blu-Ray, DVD);
 - o metadata including;
 - synopsis;
 - stars;
 - director;
 - runtime;
 - distributor;
 - genre;
 - language(s);
 - certificate;
 - year of release;
 - o links to reviews;
 - o user generated content – comments, forum posts, social media;
 - o related images and clips;
- generates income from:
 - o advertising;

¹⁵² See “About Ximon” for more details: <http://info.ximon.nl/en/about-ximon>

¹⁵³ See <http://www.findanyfilm.com/about-us>

¹⁵⁴ See <http://www.ukfilmcouncil.org.uk/>

- o affiliate fees from retail / cinema exhibitors.

Although its focus is on promoting UK film, FindAnyFilm.com contains many international films and because its trailers come from a third-party provider that licences them for any territory, does not need to show messages informing the viewer they are unavailable due to rights incompatibility.

9 Conclusions

This report has described

1. The aims of the Linked Heritage project in enabling commercial sector metadata contributions to Europeana, and the potential benefits to Europeana users and contributors;
2. Challenges and opportunities for such contributions due to the target environment (Europeana Data Model and Data Exchange Agreement);
3. A process of private sector business exploration through desk research, industry contacts and analysis of conceptual models which resulted in increased expertise with which to address these challenges and take advantage of the opportunities;
4. The business context in which cultural industries create and use product metadata, the nature of the supply chains in action, and the best practice and standards arising from this, with a comparison of the relevant underlying conceptual models;
5. A more detailed analysis of each of the four industry sectors (books, film and TV, Music and sound, photography) of interest to Linked Heritage, with their respective identifier and metadata schema standards where these exist, giving technical details, reasons for best practice and indications of coverage;
6. Some foreseen difficulties in contributing commercial product data to Europeana due to its intrinsic legal-commercial aspects;
7. Examples of public-private partnerships in several content domains which have successfully negotiated these difficulties.

There now follows a brief summary of the key findings, and recommendations based on these.

9.1 Commercial metadata in general

The central finding of this report is that commercial product metadata differs from cultural heritage object metadata in several significant ways. This is due to its origins in a business model that creates, fixes, distributes and sells mass reproductions of creative intellectual property goods, mediated by a supply chain of commercial partners bound by contract.

- Commercial datasets are not repositories of records, but flows of messages between partners:
 - o They are frequently updated depending on marketing, legal and other commercial factors (products, prices, availabilities, links & descriptors change);
 - o They must be checked and controlled for authenticity;
 - Provenance is a key issue as acting on this information has business, financial and legal consequences;
 - Identifiers therefore need to be more robust and usually involve a cost to obtain, maintain and sometimes access referent data;
- There is an intrinsic legal-commercial aspect in both individual records / messages, and in whole datasets:
 - o Given the commercial context of creating and supplying the data, there would be an implied license to re-use it for promotion of product sales;
 - o The database right in whole set of data, hence;
 - APIs;
 - licences for specific re-use;
- Metadata is costly to produce and maintain;
 - o using standards where they exist requires implementation of a compatible system, and training to enter correct details.

- o detailed datasets are seen as commercial assets;
 - they are distributed through closed data services directed only to relevant partners;
- o significant businesses depend on enriching and re-selling this data;
 - intermediary businesses have existing contracts with content creating companies¹⁵⁵;
 - data is sold to non-commercial, public sector companies.

These observations hold across the four industry sectors. However, the types of standards (and level of “standardisation”) vary considerably between the four.

9.2 Sector summary

Results by industry sub-sector follow in summarised form.

9.2.1 Cross-sector standards

Although there are interconnections between the standards, particularly for identifiers, used in the four industry sectors, there is only one truly cross-sector standard, the ISNI (International Standard Name Identifier). This is a Draft ISO standard as of 2011, though publication is not being delayed by issues with the identifier itself.

Once the ISNI standard and associated services become operative, it may offer the potential to link products and abstractions of different kinds, and across sectors, by the same creator¹⁵⁶, or provide information about people as subjects of creations.

9.2.2 Books

The books industry enjoys some of the best established identifier and metadata standards, notwithstanding some difficulties because of implementation for digital content.

- Book identifiers:
 - o In the books sector there is a well-established identifier, the ISBN, which can be used to unambiguously identify most products;
 - Publishers at present may not always apply it fully according to the standard's best practice in the case of e-books, with a possible resulting ambiguity (where a print ISBN is assigned to an e-book version, or the same ISBN to several e-book formats) or impossibility of resolution (where a non-standard number may be assigned to an e-book);
 - It has clearly defined minimum reference metadata (which can be supplied using ONIX for ISBN registration, based on ONIX 3.0);
 - o The textual abstraction identifier, the ISTC, is a new but functioning identifier and may be used to find links between products manifesting the same textual creations, and finding products that manifest related creations. It also has a set of reference metadata defined as ONIX for ISTC registration;

¹⁵⁵ As the W3C have encouragingly acknowledged: http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/#Rights_issues

¹⁵⁶ See the presentation “Smart Metadata”: <http://www.cisac.org/CisacPortal/initConsultDoc.do?idDoc=17911>

- Book metadata:
 - o The ONIX 3.0 (and the previous version, 2.1) trade standards are widely accepted and large amounts of data in this format are created by publishers. This format has a conceptual basis in indecs, and thus should map conveniently into other event-based models such as the CIDOC-CRM-derived LIDO. The ONIX format is mature and its elements enable inclusion of rich content such as subject schemes, standard identifiers such as ISBN and ISTC, and links to external content such as reviews, synopses and extracts from books.

9.2.3 Music and sound

The recorded music sector shares some features with the books industry, in that its metadata standards are also based on the indecs content model.

- Music and sound identifiers:
 - o Musical works can be identified with ISWC, a well-established, ISO standard abstraction number. It could be used to relate different recordings of a single work, e.g. recordings of various performances of a symphony. It is somewhat unclear what is the precise minimum registration data for the ISWC;
 - o The ISO-standardised ISRC identifies recordings, independently of format, and may be considered an expression identifier. Several ISRCs expressing the same musical work could be grouped with an ISWC. The ISRC has a clear set of minimum reference data. It is currently undergoing revision but this should improve its usability and maintain consistency with the current version of the standard;
 - o GRid is trade standard; an abstraction identifier for “releases”, used to group together recordings to identify the content of e.g. music albums (but independently of the medium they are manifested in, e.g. CD, digital download). Releases are manifested in concrete products e.g. a specific MP3 track for download, a CD album;
 - o For the products, GTIN-13 (formerly EAN) is the common supply chain identifier;
- Music and sound metadata:
 - o The DDEX suite of recording industry message standards includes the ERN which describes releases. This abstract entity is one level above the products actually sold to consumers, which may make it difficult to understand what is identified. The descriptive metadata offers detailed information at the individual recording level, as well as rights and marketing information.

9.2.4 Film and TV

The audiovisual sector is different from the books and music sectors in its lack of a supply chain metadata format. There are two relevant identifier standards.

- Film and TV identifiers:
 - o EIDR is a system comprising identifiers for audiovisual works, manifestations, creators, and other rightsholders, together with a DOI based metadata schema and resolution service. It has industry backing in the movie business. Its data model is compatible with indecs. Its minimum reference metadata is in the EIDR schema;
 - o The ISAN is an ISO standard number for audiovisual works. It is mainly used for televisual works and is associated with a simple registration service, but its registration data schema lacks granularity;
- Film and TV metadata:
 - o EIDR defines a comprehensive metadata schema but it is not (yet) used in the film and TV supply chain.

9.2.5 Photography

The commercial photography sector has no standard identifier, although a new system, the PLUS Registry, is proposed. The main metadata schema in use has wide recognition but is not known to be based on a documented conceptual model.

- Photography metadata:
 - o The IPTC standard for photo metadata, originally developed in the editorial photography business, is widely known and used. It covers descriptive, rights and (combined with Exif and XMP) technical details of digital photo files. However, it is normally embedded in the files themselves and it is unusual to find registries containing only image metadata. It is not apparently based on any known conceptual model, and the basis of its main descriptors in Dublin Core means it lacks semantic precision.

9.3 Public-private partnerships

The four existing commercial sector partnerships identified are in the areas of book publishing and the film and TV industry. Of these four, two face some questions with regard to long term sustainability; for Enclave (books) because the project partnership between the Spanish National Library and the Spanish Publishers' Association has now ended, new records are no longer being added and the existing links to commercial products are no longer being updated. Defining mechanisms for the continued management of such links will be key to later deliverables in this Work Package.

The original funder of FindAnyFilm.com has had its responsibilities taken over by another institution (although the website has other sources of income), and in any case its interest to Linked Heritage is mainly from the viewpoint of metadata aggregation and links to products – it does not offer access to public domain content alongside in-copyright, in-commerce items. One important aspect of FindAnyFilm's model is the use of partners who licence trailers across territorial boundaries before using them in its main discovery environment.

The remaining two partnerships identified appear to offer stable, long-term solutions to providing discovery of commercial products alongside culture heritage objects. It should be noted that in both cases the data aggregated are for products which may be very near in legal-commercial status – both out of copyright books in Gallica and archive films in Ximon.nl are likely to have originated as commercial products. It is easy to envisage users taking an interest in both historical and contemporary products in both sectors – not so easy in the music or film sectors.

All four cases show that such partnerships are in principle possible, and in some form, stable in the long term. They demonstrate that willingness to invest and participate in partnerships is certainly present in some of the countries with the largest commercial content industries markets.

However, it should be noted that this initial investment is indicative of the probable costs of setting up such partnerships to contribute to Europeana, potentially on a larger scale.

9.4 Impact

The above findings represent only the formal presentation of Task 4.1, *Private Sector Business Exploration*. Fulfilling this task has also raised the level of expertise in the Work Group, built a network of interested experts aware of the Working Group and the wider Linked Heritage project, and suggested a range of hypothetical approaches to the remaining Tasks and Deliverables of the Work Package and the overall aims of the project.

9.4.1 Contribution to Work Package 4

The work presented in this report has made a direct contribution to the progress of Work Package 4. Thanks to the knowledge synthesised here and the expertise and industry knowledge acquired, work can begin on planning and creating schema mappings and proof-of-concept data sets; enquiries can be made into suitable business models and sustainable technical solutions more clearly envisaged.

9.4.2 Relevance to other Working Groups

These results will also have relevance and interest for the other Work Groups who wish to gain an understanding of the technical, legal and commercial aspects of the metadata used in the content industries. This may be especially pertinent in the case of libraries, which, as noted in the sections on conceptual models, collect instances of commercial products and describe them in a manner oriented towards cultural heritage.

9.4.3 Broader relevance

As noted in section 5.5.2. when taking an overview of the commercial sector's metadata standards, a very limited number of projects have taken in so many domains in detail, and no known project has yet proposed attempting to integrate data from all four commercial sectors.

Given the current broad trends towards convergence through digital content, and the perennial need for identification, semantic interoperability for trading and discovery, and the potential of public-private partnerships for financial sustainability, the synthesis achieved in this report may be of interest to work in both spheres of activity. Indeed some current¹⁵⁷ and proposed¹⁵⁸ initiatives are examining new approaches to these issues.

9.5 Recommendations and Work Plan

9.5.1 Recommendations for Linked Heritage

Take-home messages for the Linked Heritage include some very positive notes, with some words of caution:

- Commercial sector metadata is rich, detailed and structured.
- There are stable identifier and metadata standards in all sectors except images:
 - o Levels of adoption and adherence vary but best practice is clearly identifiable and documented;
 - o Discovery will be well served by best practice as it is a common concern;
 - o Linking to retail opportunities can be enabled, but may need extra infrastructure and long-term commitments to manage;
 - o Conceptual models in cultural heritage and content industries are compatible;
 - Suggests mapping and aggregation will be practical and effective;
 - Suggests integration into Europeana can be successful;
 - Some extra modelling may be needed, but approaches are already suggested by existing cross-domain models.
- Public-private partnerships exist, some in long-term, stable form, across content sectors:
 - o Willingness to create partnerships is already present;
 - o Funding may be required to secure metadata feeds, access to registries, links to retail environments, supplementary marketing collateral *etc.*;

¹⁵⁷ See W3C Media Ontology: <http://www.w3.org/TR/mediaont-10/>

¹⁵⁸ See Linked Content Coalition: <http://www.linkedcontentcoalition.org/>

- o If commercial incentives exist, self-funding partnerships could aggregate data then feed into Europeana;
- Legal-commercial considerations intrinsic to the content industries' metadata flows may prevent full data sets, or data sets including full data element sets, being available for re-publication as CC0 open linked data:
 - o Aggregation of richer data could take place at a preliminary stage with full provider control over re-use rights and access, with only a selected subset of objects and/or data elements contributed to Europeana.

9.5.2 Work Plan

Fundamental understanding of the commercial sector metadata landscape has now been achieved. The understanding gained from this task has suggested some possible and probable directions to take in completing the next two deliverables.

- Further research that may be necessary (as currently envisaged) to enable aggregation and contribution to Europeana:
 - o D4.2 Technical Specification:
 - Investigatory mappings of content industry schemas to LIDO to test suitability as intermediate schema;
 - Analysis and modelling of content industry schemas to gain detailed understanding of their semantics;
 - Investigation of EDM mappings from industry sector schemas, LIDO and potentially other intermediate schema (if LIDO not suitable);
 - Feasibility study into actively managed aggregation solutions (e.g. for updating of descriptive metadata, marketing collateral or URLs to retail sites) including metadata and identifier registries, actionable identifiers and “data adaptors”¹⁵⁹;
 - Investigation into aggregators in the Europeana projects, in public-private partnerships and in commercial sector intermediary services;
 - Pilot or proof-of-concept experiments;
 - o D4.3 Legal-Commercial Framework:
 - Survey (structured or informal) of publishers' and metadata intermediaries' attitudes and concerns with regards to contributing metadata;
 - Specification of probable limits on data sets and elements aggregated, displayed at aggregation level and at Europeana discovery level;
 - Further examination of legal status of data sets and individual data objects;
 - Investigation into commercial licencing models;
 - More detailed investigation of best practice for licencing and sublicensing data flows;
 - Business models for sustainable aggregation of metadata.

It will be clear from the above considerations that many items will require synergy between the tasks leading into these two deliverables. More detailed work plans will be published to the project Website as they are developed, but the following are suggested outlines of tasks for WG4 partners:

¹⁵⁹ For the use of “data adaptor” see Peraita, 2010.

D4.2 – Technologies Chosen

Task	Partner	Deadline
Assistance with multimedia mappings Assistance with EDM / LIDO mapping	TIB	September 2012
Investigate Linked Data	NSL	September 2012
Feasibility of aggregation: metadata	MVB	September 2012
Feasibility of aggregation: identifiers	mEDRA	September 2012
Mappings UML model of ONIX Aggregation models EDM mapping Consideration of LIDO interim format	EDItEUR	September 2012

D4.3 – Legal and Licensing Framework

Task	Partner	Deadline
Advice on legal risks for libraries in PPPs	TIB	January 2013
Advice on legal risks for libraries in PPPs	NSL	January 2013
Advice on legal status of metadata repositories & services	MVB	January 2013
Advice on sustainable business model(s)	mEDRA	January 2013
Legal framework Business case(s) articulation Licencing framework(s)	EDItEUR	January 2013

Dissemination and training activity in Work Group 4 has already begun and will continue, with the following anticipated as a minimum:

- EDItEUR's routine programme of specialist meetings and events;
- Publishing sector events, such as:
 - o Frankfurt Book Fair;
 - o London Book Fair;
 - o Digital Publishing Forum¹⁶⁰.

In collaboration with other UK partners, trade events and conferences spanning the cultural heritage and academic research sectors. The other WP4 partners and WP7 are also invited to contribute.

¹⁶⁰ See <http://www.ucl.ac.uk/infostudies/digital-publishing-forum/>

10 Acknowledgements

The following people have been helpful and supportive of this report and contributing their time, expertise and in some cases personal work for inclusion.

They are listed here, alphabetically, in recognition – not to imply they are responsible for the findings or conclusions of the research.

Andreas Gebhard

Antoine Isaac

Antoine Virenque

Benoît Ginisty

Brian Green

Chris Manning

Clive Bishop

David Dawson-Pick

Detlev Balzer

FX Nuttall

Gabriele Popp

Georg Eckes

Greg Reser

Huw Morgan

J. C. Chrissey

Jeff Sedlik

Jesús Peraita

Johannes Theurer

Keith Hill

Linda Royle

Niels Rump

Peter Bloore

Peter Kilborn

Piera Costantini

Régis Flad

Sarah Saunders

Simon Cliffe

Sylvie Fodor

Valentine Charles

Yves Raimond

11 References

- Adobe Systems Inc. (2010). XMP Specification Part 1: Data model, serialization and core properties. [online]. <http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart1.pdf> [accessed August 2011].
- Albarran, A. B. (2010). *The media economy*. New York: Routledge/Taylor & Francis.
- Altomare, E. (2007). *IDEAM D3/07: Identificazione, descrizione e gestione delle immagini digitali – stato dell'arte*. [unpublished project report]. Milan:mEDRA.
- Babbage, C. (1864). *Passages from the life of a philosopher*. London: Longman, Green, Longman, Roberts, & Green.
- Bérard, R. (2011). Free library data? [online]. http://liber.library.uu.nl/publish/issues/2010-3_4/index.html?000512 [accessed July 2011]. *Liber Quarterly*. 20(3/4).
- Bell, G. (2011). *ONIX for Books: Product Information Format Implementation and Best Practice – Release 3.0*. [online] <http://www.editeur.org/93/Release-3.0-Downloads/#Best%20practice> [accessed October 2011]. London: EDItEUR.
- Bide, M. (2011a). Identifier and metadata standards for e-commerce – responding to reality in 2011. [online]. <http://dx.doi.org/10.3998/3336451.0014.108> [accessed August 2011]. *Journal of electronic publishing*. 14(1). Michigan: MPublishing.
- Bide, M. (2011b). The challenge for standards in the e-book supply chain. [online]. <http://www.editeur.org/files/Collaborations/Publications/ISQ%20Spring%202011%20-%20Mark%20Bide%20-%20Ebook-standards.pdf> [accessed July 2011]. *Information Standards Quarterly*. 23 (2). pp 25 – 30.
- Bide, M. and Bell, G. (2011). *Requirements for the identification of 'components' – an EDItEUR discussion paper*. [online]. <http://www.editeur.org/files/Collaborations/Publications/Publishers%20Requirements%20for%20Identifying%20Components%20-%20an%20EDItEUR%20discussion%20paper.pdf> [accessed July 2011].
- Bide, M. and Rust, G. (2000). *The <indecs> metadata framework. Principles, model and data dictionary*. [online]. http://www.doi.org/topics/indecs/indecs_framework_2000.pdf [accessed August 2011].
- BISG. (2010). *Roadmap of Identifiers*. [online]. http://www.bisg.org/docs/Roadmap_of_Identifiers.pdf [accessed October 2011]. New York, NY: Book Industry Study Group.
- Bloore, P. (2009). *Re-defining the Independent Film Value Chain*. [online] http://www.ukfilmcouncil.org.uk/media/pdf/h/b/Film_Value_Chain_Paper.pdf [accessed August 2011]. London: UK Film Council.
- Burnett, I. S. (2006). *The MPEG-21 book*. Chichester, England: Wiley.
- Chen, P. P-S. (1976). The Entity-Relationship Model: Toward a Unified View of Data. [online]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.123.1085> [accessed October 2011]. *ACM Transactions on Database Systems*. Vol. 1, No. 1, March.
- Crofts, N. et al.. (2010). *Definition of the CIDOC Conceptual Reference Model*. [online]. http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf [accessed September 2011]

Cooper, E. (1998). *Assess the impact of the convergence of media technologies in the home*. [online]. http://www.planetoftunes.com/research/re_media/techconv.pdf [accessed August 2011]. London, UK: PlanetOfTunes.com

Crissey, J. (2010). *Any value in the chain?* [unpublished article; online]. http://royalholloway.academia.edu/JCCrissey/Papers/646173/Any_value_in_the_chain [accessed November 2011].

Dekkers, et al. (2010). *Europeana V1.0: D3.3 Initial Technical & Logical Architecture and future work recommendations*. [online]. <http://version1.europeana.eu/web/europeana-project/technicaldocuments/> [accessed September 2011].

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].

Digital Data Exchange Standard. (2011). *Electronic Release Notification Message Suite Standard*. [unpublished]. [available upon completion of an evaluation license from <http://www.ddex.net/documents.html>].

Ding, Y. (2001). A review of ontologies with the Semantic Web in view. [online]. <http://jis.sagepub.com/content/27/6/377> [accessed August 2011]. *Journal of Information Science*. 27 (6). pp. 377-384. DOI: 10.1177/016555150102700603

Doctorow, C. *Metacrap: Putting the torch to seven straw-men of the meta-utopia*. [online]. <http://www.well.com/~doctorow/metacrap.htm> [accessed September 2011]. San Francisco, CA: The Well.

Doerr, et al. (2010). *The Europeana Data Model (EDM)*. [online]. <http://www.ifla.org/files/hq/papers/ifla76/149-doerr-en.pdf> [accessed September 2011]. Paris: International Council of Museums (CIDOC).

Dustry, I. (1999). Principle organisational structures of the music industry. [online]. http://www.planetoftunes.com/industry/industry_structure.htm [accessed August 2011]. Copyright notice as per re-use permission: © Ian Dustry. www.planetoftunes.com. 1998 – present day. UK: Planet of Tunes.

EDItEUR (2009a). *ONIX for Books product information format: Introduction to ONIX 3.0*. [online]. http://www.editeur.org/files/ONIX%203/Introduction_to_ONIX_for_Books_3.0.pdf [accessed August 2011]. London: EDItEUR.

EDItEUR. (2009b). *ONIX for ISTC registration: message format overview – version 1.0*. [online]. http://www.editeur.org/files/ONIX_for_ISTC/ONIX-ISTC_overview_v1.0.pdf [accessed August 2011]. London: EDItEUR.

EDItEUR and BNF. (2010). *State of the art and guidelines for standards applicable – Edition 2*. [online]. http://www.arrow-net.eu/sites/default/files/D4_4_State%20of%20the%20Art%20and%20guidelines_edition2.pdf [accessed August 2011].

Entertainment ID Registry Association (EIDR). (2011a). *EIDR API overview*. [online]. http://eidr.org/documents/EIDR_API_Overview.pdf [accessed October 2011].

Entertainment ID Registry Association (EIDR). (2011b). *EIDR interoperability with other standards and identifiers*. [online]. http://eidr.org/documents/EIDR_Interoperability_with_Other_Standards_Identifiers_March2011.pdf [accessed August 2011].

- Entertainment ID Registry Association (EIDR). (2010). *EIDR technical overview*. [online]. http://eidr.org/documents/EIDR_Technical_Overview_November_16_2010.pdf [accessed August 2011].
- 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> Luxembourg: Publications office of the European Union.
- Gill, T. (2004). Building semantic bridges between museums, libraries and archives: The CIDOC Conceptual Reference Model. [online]. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1145/1065> [accessed September 2011]. First Monday. 9(5-3).
- Girard, A. (2008). *Gallica 2: a global digital library*. [online]. http://www.metlib.mlp.cz/files/07_The_new_french_revolution_Aline.ppt [accessed July 2011]. Presentation at IFLA Metropolitan Libraries Section conference, Prague. June 1-6, 2008.
- Green, B. and Bide, M. (1999). *Unique Identifiers: a brief introduction*. [online]. <http://www.bic.org.uk/files/pdfs/uniqueid.pdf> [accessed November 2011]. London: BIC.
- Griffiths, S. (2011). From Dickens to Dante: ISBN propels book trade to billions. [online]. <http://www.editeur.org/files/Collaborations/Publications/ISO%20Focus%20April%202011%20-%20Stella%20Griffiths%20-%20ISBN%20Special%20Report.pdf> [accessed July 2011]. *ISO Focus*. April 2011. pp 10 – 12.
- Harris, P. R. (1991). The Development of International Standards: Exploring the ISO/ IFLA Relationship. [online]. <http://dx.doi.org/10.1177/034003529101700407> *IFLA Journal*. 17 (4). pp. 358-356.
- Hill, Keith. (1997). *CIS – a collective solution for copyright management in the digital age*. [online]. http://www.doi.org/workshop/minutes/COP_WRLD.pdf [accessed July 2011]. London: The Music Alliance.
- Holdsworth, M. (2010). *The International Standard Text Code (ISTC): A work in progress. A supply chain perspective*. [online] <http://www.bic.org.uk/files/pdfs/100401%20ISTC%20for%20BIC-BISG%20bic%20Final%20corrected.pdf> [accessed August 2011]. London: Book Industry Communications.
- IDEAlliance. (2011). *PRISM Image Metadata Specification*. [online]. <http://www.idealliance.org/downloads/prism-image-metadata-specification> [accessed August 2011]. Alexandria, VA: IDEAlliance.
- IFFRO. (2010). *How to map ONIX For RROs to your system*. [online]. http://www.editeur.org/files/ONIX%20for%20RROS/ONIX_for_RROs_Handbook.pdf [accessed October 2011]. London: EDItEUR.
- IFLA. (2009). *Functional requirements for bibliographic databases: Final report*. [online]. http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf [accessed August 2011].
- IFPI. (2007a). *Global Release Identifier (GRid) standard: version 2.1*. [online]. www.ifpi.org/content/library/GRid_Standard_v2_1.pdf [accessed July 2011].
- IFPI. (2007b). *The GRid handbook: a users' guide to the implementation of the Global Release Identifier (GRid) – version 2.0*. [online]. [accessed July 2011].
- IMCA, (2002?). *Identification and evaluation of financial flows within the European cinema industry by comparison with the American model*. [online]

http://ec.europa.eu/avpolicy/docs/library/studies/finalised/film_rating/sum_en.pdf [accessed August 2011]. Study n° DG EAC/34/01. Brussels: European Commission.

International DOI Foundation. (2003). *DOI news: 18 March 2002*. [online].

<http://www.doi.org/news/020319-Editeur.html> [accessed August 2011]. Oxford: Internat. DOI Foundation.

International DOI Foundation. (2011). *DOI Handbook*. [online].

http://www.doi.org/handbook_2000/metadata.html [accessed August 2011]. Oxford, UK: IDF.

International ISAN Agency. (2005). *ISAN user guide*. [online].

http://www.isan.org/docs/ISAN_User_Guide_2.2.pdf [accessed July 2011]. Geneva: ISAN Internat. Agency.

International ISMN Agency. (2008). *ISMN user's manual*. [online]. [http://www.ismn-](http://www.ismn-international.org/download/Web_ISMN_Users_Manual_2008-4.pdf)

[international.org/download/Web_ISMN_Users_Manual_2008-4.pdf](http://www.ismn-international.org/download/Web_ISMN_Users_Manual_2008-4.pdf) [accessed August 2011]. Berlin: Internat. ISMN Agency.

International ISNI Agency. (2009). *ISNI Frequently Asked Questions. Version 2.0*. [online].

http://www.isni.org/docs/isni_faq.pdf [accessed July 2011].

International ISBN Agency. (2009). *International ISBN agency*. [online]. [http://www.isbn-](http://www.isbn-international.org/page/about)

[international.org/page/about](http://www.isbn-international.org/page/about) [accessed July 2011] London: Internat. ISBN Agency.

International ISBN Agency. (2010). *E-books and ISBNs: a position paper and action points from the International ISBN Agency*. [online]. http://www.isbn-international.org/pages/media/ISBN_e-book_paper_Feb2010.pdf [accessed August 2011].

International ISRC Agency. (2009). *International Standard Recording Code (ISRC) Handbook: 3rd edition*. [online].

http://www.ifpi.org/content/library/isrc_handbook.pdf [accessed August 2011]. London, UK: IFPI.

International ISTC Agency. (2010). *International Standard Text Code (ISTC) User Manual*. [online]

http://www.istc-international.org/html/multimedia/pdfs/ISTC_User_Manual_2010v1.2.pdf [accessed July 2011]. London: Internat. ISTC Agency.

IPTC. (2010). *IPTC standard photo metadata: IPTC Core specification 1.1 / IPTC Extension*

specification 1.1. [online]. http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata-201007_1.pdf [accessed September 2011].

Isaac, A. et al. (2010). *Europeana Data Model Primer*. [online].

<http://version1.europeana.eu/web/europeana-project/technicaldocuments/> [accessed September 2011].

ISO. (2011). *ISO/IEC Directives, Part 2, Rules for the structure and drafting of International Standards*.

[online]. <http://isotc.iso.org/livelink/livelink?func=ll&objId=4230456> [accessed October 2011]. 6th ed. Geneva: ISO.

ISO/IEC. (2004). *Information technology — Metadata registries (MDR) — Part 1: Framework*. [online].

<http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html> [accessed October 2011]. Geneva: ISO.

JEITA (Japanese Electronics and Information Technology Industries Association). (2010). *JEITA Standards*. [online].

http://www.jeita.or.jp/english/standard/html/1_4.html [accessed August 2010]. Tokyo: JEITA.

JISC. (2006). *CLIC Project Final Report*. [online].

http://www.jisc.ac.uk/uploaded_documents/CLIC_Report.pdf [accessed September 2011].

- Johnson, E. (2011). Intellectual Property and the incentive fallacy. http://www.law.depaul.edu/centers_institutes/ciplit/ipsc2011/pdf/Papers/JohnsonE_Paper.pdf [accessed August 2011]. *Florida State University Law Review*, forthcoming. Chicago, Illinois: De Paul College of Law. July 25th, 2011.
- KEA European Affairs. (2006). *The economy of culture in Europe*. [online]. <http://www.keanet.eu/ecoculture/studynew.pdf> [accessed August 2011].
- Lagoze, C., Hunter, J. and Brickley, D. (2000). An event-aware model for metadata interoperability. [online]. <http://itee.uq.edu.au/~ereseach/papers/2000/harmonyECDL2000.pdf> [accessed September 2011].
- Leach, P. J. (2005). *A Universally Unique Identifier (UUID) URN Namespace*. [online] <http://www.ietf.org/rfc/rfc4122.txt> [accessed August 2011]. Fremont, California: Internet Engineering Task Force.
- Linked Heritage (2011). About [online] <http://www.linkedheritage.org/index.php?en/138/about> [accessed July 2011].
- Luther, J. (2009). *Streamlining book metadata workflow*. [online]. http://www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf [accessed August 2011]. Baltimore, MD: NISO; Dublin, Ohio: OCLC.
- McKenna, G. and De Loof, C. (2009). *Report on existing standards applied by European museums*. [online]. <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents> [accessed July 2011].
- McKenna, G. and Wyns, R. (2010). *Assessment of requirements for persistent identification of objects, collections and institutions*. [online] <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents> [accessed July 2011].
- mEDRA (2002a). *DOI market survey report*. [online] doi: http://dx.doi.org/10.1392/market_survey [accessed July 2011].
- mEDRA. (2002b). *Media-related identification and metadata standards*. [online]. http://www.medra.org/stdoc/md_standards.pdf [accessed August 2011]. doi: 10.1392/metadata_standards
- MWG (Metadata Working Group). (2010). *Guidelines for handling metadata: version 2.0*. [online]. http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf [accessed July 2011].
- Nash Information Services, LLC. (2011). *Market share for each distributor in 2011*. [online]. <http://www.the-numbers.com/market/Distributors2011.php> [accessed August 2011]. Beverly Hills, CA: Nash Information Services, LLC.
- NISO. (2004). *Understanding Metadata*. [online]. www.niso.org/publications/press/UnderstandingMetadata.pdf [accessed September 2011]. NISO Press.
- ODL. (2005). *Metadata in the Oxford Digital Library*. [online]. <http://www.odl.ox.ac.uk/metadata.htm> [accessed August 2011].
- Open Archives Initiative. (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*. [online]. <http://www.openarchives.org/OAI/openarchivesprotocol.html> [accessed September 2011].
- Park, J. R. and Childress, E. (2009). Dublin Core metadata semantics: an analysis of the perspectives of information professionals. [online]. <http://dx.doi.org/10.1177/0165551509337871> [accessed August 2011]. *Journal of Information Science* 35 (6) pp. 727-739.

- Peraita, J. (2010). *Spain's DILVE experience: a BIP repository as data adaptor for publishers*. [online]. http://www.editeur.org/files/Events%20pdfs/Supply%20chain%202010/Peraita_iscs_frankfurt2010_dilve_v02_nt.pdf [accessed July 2011]. Presentation given at 32nd International Supply Chain Seminar, Frankfurt. 5 October 2010.
- Rust, G. (1996). *The common information system (CIS)*. [online]. <http://www.cisac.org/CisacPortal/documentLink.do?sw=on&lang=en&id=17142> [accessed August 2011]. Paris: CISAC.
- 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> [accessed August 2011].
- Rust, G. (2005a). *The model of making in indecs and RDD*. [online]. <http://www.doi.org/topics/RustModelofMaking2005.pdf> [accessed October 2011].
- Rust, G. (2005b). *Thoughts from a different planet*. [online]. http://www.oclc.org/research/activities/past/orprojects/frbr/frbr-workshop/presentations/rust/050502_Godfrey_Rust_FRBR_presentation.ppt [accessed August 2011]. Presentation at FRBR Workshop, Ohio May 4 2005. Dublin, Ohio: OCLC.
- Saunders, S. and IPTC Photometadata Working Group. *Image Metadata Handbook*. [online]. http://www.cepic.org/issues/2011/06/image_metadata_handbook [accessed October 2011]. Berlin: CEPIC.
- Simmonds, A. W. (1999). Identifiers. *Publishing Research Quarterly*. Summer 1999. 15 (2).
- Sleep, T. (2011). On the PLUS side. [online]. <http://www.epuk.org/Opinion/985/on-the-plus-side> [accessed September 2011]. *EPUK Opinion*. 25 July 2011. London, UK: Editorial Photographers UK Ltd.
- Sowa, J. (2000). Ontology, metadata and semiotics. [online]. <http://www.jfsowa.com/ontology/ontometa.htm> [accessed August 2011]. Also published in: *Lecture Notes in Computer Science*. 1867. pp. 55-81. DOI: 10.1007/10722280_5
- STOA (Scientific and Technological Options Assessment) Study Panel. (2001). *Globalisation of the media industry and possible threats to cultural diversity*. [online]. http://www.europarl.europa.eu/stoa/publications/studies/20001301_en.pdf [accessed August 2011]. Brussels: European Parliament.
- Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. Cambridge, Mass: MIT Press.
- Taylor, A. G. (2004). *The organization of information*. Westport, Conn: Libraries Unlimited.
- Tsuru Zoh, T. (2001). *Description of Exif file format*. [online]. <http://web.archive.org/web/20110716154113/http://park2.wakwak.com/~tsuruzoh/Computer/Digicams/exif-e.html> [accessed August 2011].
- Vaisberg, L. (2010). [Response of the FEP to] *Green paper on "Unlocking the potential of cultural and creative industries"*. [online] http://ec.europa.eu/culture/our-policy-development/consultation/EU_organisations/A-134-O-Federation_European_Publishers_FEP.pdf [accessed July 2011].

12 Appendix 1 – Glossary of terms

Term	Explanation	Sample definitions and further notes	SEE ALSO:	Original context
Abstraction	Concept derived from perceptible instances or examples Key term in indecs because it is the general class of intellectual property, the entity traded in the content or copyright industries CIDOC equivalent: "E28 Conceptual object"	"A creation which is a concept; an abstract creation whose existence and nature are inferred from one or more expressions or manifestations" http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "The "abstracted" work, according to any party who has reason to identify a conceptual "work" underlying a particular manifestation." http://www.doi.org/topics/RustModelofMaking2005.pdf NOTE: more general than Functional Requirements for Bibliographic Records (FRBR) terms "work", "expression" and "manifestation" - can encompass all three	Aggregation, Cultural heritage object, Intellectual property, Work, Manifestation	INDECS
Actionable / actionability	When identifiers are associated with services (normally resolution)	"Of identifiers...: implying that in the context of a particular piece of digital network infrastructure, the identifier can be readily used to perform some action: e.g. in an Internet Web browser, it can be "clicked on" and some action takes place." http://www.doi.org/handbook_2000/glossary.html	Resolution, Service	DOI ISBN-A
Aggregated resource	One of the digital, physical or conceptual "things" in an aggregation	"An Aggregated Resource is a Resource that is a constituent of an Aggregation due to an assertion in a Resource Map that describes the containing Aggregation." http://www.openarchives.org/ore/1.0/datamodel#Aggregated_Resource	Aggregation, Entity, Resource map, Resource	OPEN ARCHIVES

Aggregation	Entity used in Europeana Semantic Elements to represent cultural heritage objects (CHOs), required because various resources are used to give access to different aspects of the CHO during Europeana's archival process A "conceptual construct", so an "abstraction" in the Index terminology	"An Aggregation is a Resource of type ore:Aggregation that is a set of other Resources... an Aggregation is a conceptual construct" http://www.openarchives.org/ore/1.0/datamodel#Aggregation "A set of related resources (Aggregated Resources), grouped together such that the set can be treated as a single resource. This is the entity described within the ORE interoperability framework by a Resource Map" http://version1.europeana.eu/c/document_library/get_file?uuid=7ef62859-4846-4697-9285-6bde99d976f5&groupId=10605	Entity, Europeana aggregation, Abstraction, Resource, Type	OPEN ARCHIVES EUROPEANA
Application profile	Used in both DOI and Dublin Core (DC) domains. Includes a metadata element specification that in DC can include modifications to element definitions	"A set of metadata elements, policies, and guidelines defined for a particular application. The elements may be from one or more element sets, thus allowing a given application to meet its functional requirements by using metadata from several element sets including locally defined sets. For example, a given application might choose a subset of the Dublin Core that meets its needs, or may include elements from the Dublin Core, another element set, and several locally defined elements, all combined in a single schema. An Application profile is not complete without documentation that defines the policies and best practices appropriate to the application." http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml#A "the functional specification of the application profile includes a set of metadata, comprising the kernel metadata and additional information applicable to that particular genre of object and functional requirements." http://www.doi.org/handbook_2000/glossary.html#ap	Governance, Schema, Service	DOI DUBLIN CORE

Attribute	A generalised characteristic of all entities in Indexes In the XML syntax, adds metadata for specification to a data element itself	“A characteristic of an entity (adapted from ISO 11179); something which an entity has” http://www.doi.org/topics/indexes/indexes_framework_2000.pdf “The Name-AttValue pairs are referred to as the attribute specifications of the element... with the Name in each pair referred to as the attribute name...and... the content of the AttValue (the text between the ' or " delimiters) as the attribute value” http://www.w3.org/TR/REC-xml/#dt-attr	Entity, Data model	INDEXS XML
Books in print	Database of all published (printed) books available for supply from a publisher (who decides to make the product available) to retailers	[these types of product databases not covered in detail by this report but will be discussed in future deliverables]	In print, In commerce	ARROW
Cardinality	Range of possible numbers of a data element for a given entity	“The notation (Min, Max) ... indicates the minimum number of times an element MUST occur, and the maximum number of times an element MAY occur. The notation used in the ONIX documentation follows the format "n – n". This attribute of a data element is sometimes also referred to as its cardinality.” http://www.editeur.org/files/ONIX%20for%20RROS/ONIX_for_RROs_Handbook.pdf	Attribute, Conceptual model, Data element, Entity	ONIX
Choreography (“message choreography”)	Description of all possible (or all most likely) interchanges of (metadata) messages between two or more distinct parties involved in the supply chain within one media sector	“A choreography defines re-usable common rules that govern the ordering of exchanged messages, and the provisioning patterns of collaborative behaviour, as agreed upon between two or more interacting participants” http://www.w3.org/TR/ws-cdl-10/#Choreography “By “choreography”, we mean the rules by which messages are exchanged” http://www.editeur.org/files/ONIX%20for%20ISTC/ONIX-ISTC%20overview%20v1.0.pdf	Record	DDEX ONIX W3C

Codelist	Codelists are controlled vocabularies where the preferred terms are language-independent "codes". Benefits: language independence and reduced message length for machine-to-machine use	"ONIX uses codes to provide a measure of explicit language-independence. Codelists consist of tables with three main columns..." [notation (code), label, and explanatory (scope) note] http://www.editeur.org/93/Release-3.0-Downloads/#Best%20practice	Controlled vocabulary, Value domain	ONIX
Commerce / commercial	Transactions by which goods are exchanged In this report, mainly financial transactions for creative content and products	"People make stuff. People use stuff. People do deals about stuff. The stuff and the deals may come in any order, but neither come before the people. This is the basic model of commerce that underlies the <indecs> framework and models... Commerce is used here in its broadest sense, not necessarily having financial gain as its object" http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "transactions (sales and purchases) having the objective of supplying commodities (goods and services)" http://wordnetweb.princeton.edu/perl/webwn?s=commerce	Commercial sector, Content industries, In commerce	INDECS

Commercial sector / purposes / organisation	The commercial sector is distinguished from the public sector in this report by the nature of its activity: creating cultural objects which are protected by copyright and (can) produce financial reward Photography sometimes incorrectly placed in “non-industrial sectors” but is included for this report as it belongs to the “content industries”	“Industrial sectors producing cultural products aimed at mass reproduction, mass-dissemination and exports (for example, a book, a film, a sound recording). These are “cultural industries” including film and video... broadcasting, music, book and press publishing.” http://www.keanet.eu/ecoculture/studynew.pdf	Commerce / commercial, Content industries	KEA
Composite (element)	An element with one or more sub-elements within it to specify several attributes of the same entity in one statement, or make structured statements	“A composite data element contains two or more component data elements.” http://www.editeur.org/files/EDIfact%20eancom%20pdfs/EDIfact%20trade%20book%20supply/T2v11%20%28trade%20books%20supply%20-%20structure%29.pdf	Element, Data model, Data structure	EDIFACT

Concept (“conceptual object”)	One of the two top classifications of entity in Indecs, as opposed to percept In CIDOC the equivalent “conceptual object” is identified but the need for a physical medium (even if “only” a human memory) is explicit – in Indecs this dependence is implicit	“An entity which cannot be perceived directly through the mode of one of the five senses... an abstract entity, a notion or idea; an abstract noun; an unobservable proposition which exists independently of time and space” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “This class comprises non-material products of our minds and other human produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication... Conceptual Object[s] have the ability to exist on more than one particular carrier at the same time, such as paper, electronic signals, marks, audio media, paintings, photos, human memories, <i>etc...</i> They cannot be destroyed. They exist as long as they can be found on at least one carrier or in at least one human memory. Their existence ends when the last carrier and the last memory are lost.” http://www.cidoc-crm.org/official_release_cidoc.html	Abstraction, Percept	INDECS CIDOC-CRM
Conceptual model	A model of possible entities and relationships to be represented by a metadata schema, based on domain knowledge	[Conceptual models include the three referred to in this report: indecs,CIDOC-CRM and FRBR(oo)]	Data model, Data dictionary	

Content	Europeana: metadata, including previews and links, that “represent” cultural heritage objects, by means of aggregation of resources representing CHOs Cultural industries (indecs): intellectual property; also refers for convenience, to the manifestations from which abstractions (conceptual creations) are derived	“Europeana stores representations of digital objects and not the digital objects themselves... Europeana data includes: Metadata (descriptive, administrative) describing a digital object; Preview (thumbnail) of the described object; Active links to the described digital object on the provider’s site” http://www.version1.europeana.eu/web/europeana-project/ec-information-day “Europeana provides integrated access to digital objects from the cultural heritage organisations of all the nations of the European Union... To do this Europeana harvests and indexes the descriptive metadata associated with the digital objects.” http://www.version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&groupId=10602 “An abstraction is the entity often popularly called a work. However, in the <indecs> framework a work is a piece of intellectual property (ip) defined directly in terms of the legal provisions of the Berne Convention, so while all works are abstractions, all abstractions are not necessarily works in the legal sense... Common formal elements, such as a storyline, or a melody, or the arrangements of words and images, are typical evidence of a common underlying abstraction, and it is these elements of expression which are subject to copyright; but the precise characteristics by which such recognition is secured are elusive and are settled by editorial, commercial or, ultimately, by a legal judgement.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Aggregation, Aggregated resource, Content industries, Creation, Intellectual property, Manifestation	EUROPEANA INDECS
Content industries	The cultural industries identified in this report, which rely on intellectual property protected by copyright	“...the major content sectors of text, music, images and audiovisual...” http://www.doi.org/topics/RustModelofMaking2005.pdf [this definition is to be preferred to that referenced in “Cultural industries”]	Abstraction, Content, Cultural industries, Creation, Intellectual property	INDECS

Controlled vocabulary	A set of allowable content values, specified by a standard and/or a named governance organisation, for a given set of metadata elements The simplest way to validate data (for example formulae could be used instead of fixed values)	“Metadata schemes... may specify content rules for how content must be formulated (for example...), allowable content values (for example, terms must be used from a specified controlled vocabulary).” www.niso.org/publications/press/UnderstandingMetadata.pdf	Codelist, Data element, Value domain	NISO
Creation	In Indecs one of the key categories grouping together the concepts describing the process and outputs of the creative industries	“The output of creative activity” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Abstraction, Expression, Manifestation	INDECS

Cultural heritage object	For this report's purpose, the public domain materials curated by publicly-funded institutions such as museums, archives, galleries and libraries comprise cultural heritage Culture is a “process or performance”, and thus a series of “events” in the Index and CIDOC models. These events generate abstract concepts, which are then expressed in manifestation singletons called “cultural heritage objects”	“Heritage is a cultural process or performance that is concerned with the types of production and the negotiation of cultural identity, individual and collective memory, and social and cultural values” Quoted in: http://icom.museum/what-we-do/professional-standards/key-concepts-of-museology.html	Cultural industries, Content, Europeana aggregation	ICOM
Cultural industries	Industries characterised by: Essentially intellectual creations protected by copyright law; intellectual property sold or licensed in mass reproduction For the purposes of this report the “cultural industries” are identical with “the commercial sector”	“Industrial sectors producing cultural products aimed at mass reproduction, mass-dissemination and exports (for example, a book, a film, a sound recording). These are “cultural industries” including film and video... broadcasting, music, book and press publishing.” http://www.keanet.eu/ecoculture/studynew.pdf [this process-centred definition is incomplete compared to that given for “Content industries”, and the latter is to be preferred]	Content industries, Commercial sector, Cultural heritage	KEA

Data dictionary	An ontology for a data model or framework, usually expressed in technical but human-readable terms, layout and syntax for users' understanding	"A structured ontology, developed from the indecs Framework, containing all Terms used in DOI® AP Metadata Declarations, ONIX messages and other schemes, and formal mappings of the relationships between them." http://www.doi.org/handbook_2000/glossary.html#ap "The framework metadata dictionary holds information on <indecs> metadata elements, their names, iids, definitions, relationships and mappings to elements in other schemas." http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Data model	INDECS DOI
Data element	The smallest unit of a data model, schema or structure; the "container" which holds the smallest item of data used in the model or application. Often used to refer to the actual metadata that is found in a given slot	"An item of metadata" http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "...the basic container for data is called a data element. It may exist purely as an abstraction or exist in some application system" ... "unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes" http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html	Field, Data model, Data structure	ISO/IEC 11179
Data model ("conceptual [data] model")	A conceptual model that lays out what can be described, and how, by a given metadata (standard) framework Either explicitly or implicitly based on some analysis of the domain of knowledge or practice it serves	"The... model elaborates a logical and semantic framework for describing entities, their attributes and, where appropriate, values of each. Entities, attributes and values are referred to as types of metadata elements." http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "graphical and/or lexical representation of data, specifying their properties, structure and inter-relationships" ... "conceptual data model / conceptual model: data model that represents an abstract view of the real world. NOTE: A conceptual model represents the human understanding of a system." http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html	Conceptual model, Data dictionary, Data structure, Schema	ISO/IEC 11179

Data structure	The concrete way a specific set of data elements are related; either in a discrete metadata message or record, or stored in a fixed database system	“A hierarchical description of a set of data elements. A data set can be described according to its data structure.” http://www.w3.org/2003/glossary/alpha/D/	Schema, Data dictionary, Data model	W3C
Documentation	Information provided for specific users, that can be incorporated into metadata Here, it applies to the technical, academic and practitioner level information created in the cultural heritage sector	“identifiable immaterial items that make propositions about reality... Documentation databases are regarded as a special case of E31 Document.” ... “The term “scientific documentation” is intended to convey the requirement that the depth and quality of descriptive information that can be handled by the CRM should be sufficient for serious academic research. This does not mean that information intended for presentation to members of the general public is excluded, but rather that the CRM is intended to provide the level of detail and precision expected and required by museum professionals and researchers in the field.” http://www.cidoc-crm.org/official_release_cidoc.html	Cultural heritage object, Record	CIDOC-CRM
Edition	An attribute at the expression level in FRBR, but corresponding to the manifestation level in Indexes – this term is included to help show the differences in approach between the two models, the relatively “flat” FRBR and the event-based Indexes	“A distinguishing characteristic... of the expression... serves to differentiate the expression from another expression of the same work (e.g., ...an “edition” or version designation” http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf “All copies of a book, pamphlet, fascicle, single sheet, etc., printed from the same typographic image and issued by the same entity in the same format at one time or at intervals without alteration. An edition may consist of several impressions in which the text and other matter are not substantially changed... Also applies to one of the formats in which a literary work or collection of works is published, usually for a specific purpose or market, for example, a book club edition, colonial edition, deluxe edition, export edition, library edition, limited edition, paperback edition, shorthand edition, special edition, or trade edition. It is not uncommon for a new book to be published in multiple editions...” http://www.abc-clio.com/ODLIS/odlis_e.aspx#edition	Expression, Manifestation	FRBR LIBRARY SCIENCE

Embedded metadata / content	Primarily refers to photo metadata; including the metadata about a digital file within the file itself; digital files containing creations can also be embedded in metadata e.g. for marketing purposes	“The two main places to enter metadata are to embed it in the image file (as Exif, IPTC and PLUS) and to store it outside the image file in a database.” http://www.cepic.org/issues/2011/06/image_metadata_handbook	Content, Data element, Manifestation, Record, Resource	CEPIC IPTC
Entity	In the Indecs framework, entities are identified objects of interest, whether conceptual or directly perceived by the senses	“an entity is something which is identified” ... “a thing must be both thought about or perceived and identified before it exists in a metadata framework. This is more like the term resource in the sense adopted by the World Wide Web Consortium...” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Identifier, Resource, Referent	INDECS W3C
Europeana aggregation	“This class is used in Europeana to gather in a single conceptual unit all the information about a cultural heritage object, necessary for all operations on these objects.” (EDM specification)	“The set of resources related to a single cultural heritage object that collectively represent that object in Europeana. Such set consists of: all descriptions about the object that Europeana collects from (possibly different) content providers, including thumbnails and other forms of abstractions, as well as of the description of the object Europeana builds.” http://version1.europeana.eu/c/document_library/get_file?uuid=7ef62859-4846-4697-9285-6bde99d976f5&groupId=10605	Aggregation, Aggregated resource, Content, Cultural heritage object	EUROPEANA

Expression	Event in time and space that makes creative ideas perceptible, e.g. a musical work has an expression in a concert which is manifested in many ways, including recordings of the concert In contrast, in FRBR the term denotes an abstraction (in indecs terms) defined by common formal symbolic structures	“A creation which is an event... An expression is a performance – an event which is in itself regarded as a creation and may or may not be recorded or fixed in a manifestation, and may be reproduced by some form of recording playback. The expression is the event which is recorded, not the physical or digital recording itself, which is a manifestation. One expression may be recorded and copied onto many media while maintaining its integrity. An expression may give rise to an abstract work; at the same time it may be an expression of an existing abstract work.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “the intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, <i>etc.</i>, or any combination of such forms... Expression encompasses, for example, the specific words, sentences, paragraphs, <i>etc.</i> that result from the realization of a work in the form of a text, or the particular sounds, phrasing, <i>etc.</i> resulting from the realization of a musical work. The boundaries of the entity expression are defined, however, so as to exclude aspects of physical form, such as typeface and page layout, that are not integral to the intellectual or artistic realization of the work as such.” http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf	Abstraction, Data model, Edition, Entity	INDECS FRBR
Format	The format of an information object carrier refers to the physical characteristics that enable symbols to be stored there, before any actual storage takes place	“A format is an artefact which requires content to become a manifestation: a blank DVD, an empty computer file or a book without words.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “StorageMediumFormat: The physical form of the material on which the content of the resource is stored. HousingFormat: The physical format of the encasing for the storage medium. Note: Applies to resources conveying content on a physical carrier. Resources conveying content by means of broadcasting or online transmission will have a HousingFormat value not applicable.” http://www.loc.gov/marc/marbi/2007/5chair10.pdf	Manifestation, Media	INDECS RDA /ONIX

Governance	Governance is the rule-making activity (sometimes including also maintenance, development and technical support) implied by any identifier and/or metadata framework (and service). The DOI system is used here as an example of all the above aspects	“the governance body... safeguards (owns or licences on behalf of registrants) all intellectual property rights relating to the DOI System. It works with RAs and with the underlying technical standards of the DOI System components to ensure that any improvements made to the system (including creation, maintenance, registration, resolution and policymaking of DOI names) are available to any DOI name registrant, and that no third party licenses might reasonably be required to practice the DOI name standard. DOI name resolution is freely available to any user encountering a DOI name... [it] provides implementation through agreed standards of governance and scope, policy, to define “rules of the road”. It also provides a technical infrastructure (resolution mechanism, proxy servers, mirrors, back-up, central dictionary) and a social infrastructure (persistence commitments, fall-back procedures, cost-recovery (on a self-sustaining model), and shared use of the system. The IDF is not a standards body, but a central authority and maintenance agency” http://www.doi.org/overview/sys_overview_021601.html	Registration, Service	DOI
Group	A secondary division within ONIX for Books “blocks”; a group of metadata elements with a similar purpose and meaning	“Each block consists of one or more numbered “data element groups”. Data element groups are defined primarily for convenience in presentation of the detailed record specification. They do not necessarily have any structural significance: some are single composites, and some are flat sequences of elements” http://www.editeur.org/files/ONIX%203/Introduction_to_ONIX_for_Books_3.0.pdf	Data structure, Header, Record	ONIX
Header	Part of a message or record which identifies the type of message or record format used, often contains information about the sender / creator, the time and date, etc. ONIX header described here as illustrative of typical best practice	“must specify the sender of the message...an organization name, plus any applicable organization identifiers; details for a technical contact able to deal with queries about the message should be included; should specify the recipient(s) of the message... when the ONIX message is tailored for a specific addressee, for example through a specific selection of products that would not be suitable for other recipients; should specify message sequence number when the ONIX message is a ‘delta file’; must specify the date the message was sent” http://www.editeur.org/93/Release-3.0-Downloads/#Best%20practice	Data structure, Group, Record	ONIX DDEX

Identifier	A unique name or label given to a resource or entity In this report, a standard unique identifier in numerical or alphanumerical code form	“A unique label allocated to an entity within a given namespace” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “An unambiguous string or “label” that references an entity” http://www.doi.org/handbook_2000/glossary.html	Entity, Intelligent (identifier), Namespace, Resource, Service	INDECS DOI
In commerce	In-copyright products that are also available for sale, defined by opposition to out-of-commerce in the sample definition here (the significant factor being the decision to publish and sell)	“Out-of-commerce works are works that are still protected by copyright but are no longer commercially available because the authors and publishers have decided neither to publish new editions nor to sell copies through the customary channels of commerce. In the past works such as books were referred to as being either “in-print” or “out-of-print”. Today, with the advent of electronic channels of commerce, the term “out-of-commerce” is used (with electronic publishing a book will be “in commerce” even if only available in electronic form).” http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/619	In print	EU
In print	Available for trading (not necessarily in copyright)	[details of the book publication and releasing processes of other content industries not discussed in detail here, but will be addressed in future deliverables]	Books in print, In commerce	
Intellectual property	A legal concept depending on the particular jurisdiction in question, governing ownership and use of content	“An entity defined by law or international convention to be intellectual property” [or more generally] “A concept defined by law, statute or international convention” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Content	INDECS

Intelligent (identifier)	The most intelligent kind of identifier is a name or title which unambiguously identifies the resource in the right context; codes and serial numbers may be directly humanly resolvable too, (e.g. "Agent 007") but the more removed from context they are, the less "intelligent" they are, and the fewer connotations they bring	"identifiers which carry some information in their structure relating to the entity they identify, such as a format, date or producer code" http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "This is probably best described by reference to what an unintelligent numbering scheme is: this is a purely random number which can only be interpreted by reference to a central database; examining the number itself tells you nothing about the object which it identifies." http://www.bic.org.uk/files/pdfs/uniqueid.pdf	Identifier	INDECS
Linked (Open) Data	The approach used to design a "Web of data" (as opposed to separate documents) which can be browsed, machine-read to create search results and logical inferences, and combined together more easily into meaningful aggregations In metadata terms, data elements are linked, and are especially useful if they contain identifiers (so linked data <i>per se</i> is not necessarily openly accessible)	"Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they [are] links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow: Use URIs as names for things / Use HTTP URIs so that people can look up those names. / When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL) / Include links to other URIs. so that they can discover more things." http://www.w3.org/DesignIssues/LinkedData "The Semantic Web is a Web of Data — of dates and titles and part numbers and chemical properties and any other data one might conceive of. The collection of Semantic Web technologies (RDF, OWL, SKOS, SPARQL, etc.) provides an environment where application can query that data, draw inferences using vocabularies, etc... not only does the Semantic Web need access to data, but relationships among data should be made available, too, to create a Web of Data (as opposed to a sheer collection of datasets). This collection of interrelated datasets on the Web can also be referred to as Linked Data." http://www.w3.org/standards/semanticweb/data	Aggregation, Data element, Identifier, RDF	W3C

Manifest	A list of the contents of an electronic or analogue information package, separately accessible from the information package itself	“For large data files received via FTP as part of an electronic feed, a “manifest” is needed to identify the contents to save staff the time of having to open the file to learn what is in it.” http://www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf	Content, Aggregation Metadata, Product, Release	NISO
Manifestation	Set of all perceptible embodiments of a creative work in one specific format and medium; in practice somewhat abstract, because one is unlikely to experience all identical products together in any meaningful way This term has the same meaning in both FRBR and Indecs, and was added to CIDOC-CRM by the FRBR(oo) harmonisation	“An artefact containing an inflexion of an expression” ... “a manifestation is typically not an individual creation but a class of creations. For example, in describing a book with its ISBN, format, title, author and subject classes, we are describing all instances of that book.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form. When a work is realized, the resulting expression of the work may be physically embodied on or in a medium such as paper, audio tape, video tape, canvas, plaster, etc. That physical embodiment constitutes a manifestation of the work.” http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf	Abstraction, Edition, Expression	INDECS FRBR
Media	Media are types of physical artefacts for transmission and storage of symbols that stand for conceptual creations (in Indecs, this is called “format” but common use of the word includes aspects of Indecs “genre” too)	“A format is an artefact which requires content to become a manifestation: a blank DVD, an empty computer file or a book without words.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “genre: A style or manner of the expression of an abstraction (lexical, musical, pictorial, audiovisual, narrative etc.)” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Format, Manifestation	RDA-ONIX INDECS

Metadata	Metadata (plural) are commonly defined as “data about data” (see NISO, 2004) but are given a specific technical meaning in this report, from both a commercial and Linked Data point of view The technical meaning “a relationship someone claims to exist between two entities” here highlights the role of the asserter of the claim, not just the assertion	“...an item of metadata is a relationship that someone claims exists between two entities” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf [cited in: http://www.doi.org/handbook_2000/resolution.html] “Literally, “data about data.” Structured information describing information resources/objects for a variety of purposes... Metadata has been categorized as descriptive, structural, and administrative. Descriptive metadata facilitates indexing, discovery, identification, and selection. Structural metadata describes the internal structure of complex information resources. Administrative metadata aids in the management of resources and may include rights management metadata, preservation metadata, and technical metadata describing the physical characteristics of a resource.” http://www.abc-clio.com/ODLIS/odlis_m.aspx#metadata	Attribute, Data element, Data structure, Entity, Party, Relation	INDECS LIBRARY SCIENCE
Monograph (including publication)	A more precise term for the products of the book publishing industry, focussed more on the intellectual content and publication process than the physical format and medium	“any nonserial publication, complete in one volume or intended to be completed in a finite number of parts issued at regular or irregular intervals, containing a single work or collection of works. Monographs are sometimes published in monographic series and subseries” http://www.abc-clio.com/ODLIS/odlis_m.aspx#monograph	Edition, In print, In commerce, Manifestation	LIBRARY SCIENCE
Multiple resolution	Providing information and/or resources related to an identifier's referent and relevant to the context of the enquiry	“Resolution returning a specific output of several pieces of current information related to the DOI name identified entity, specifically at least one URL plus defined data structures representing Application Profile and Service information. Multiple resolution goes hand-in-hand with managing the object and its available services through structured metadata” http://www.doi.org/handbook_2000/resolution.html#3.3	Actionability, Identifier, Resolution, Service	DOI ISBN-A

Namespace	Set of elements available in a given schema, corresponding to one domain of knowledge or practice. The domain giving the context of the terms must be identified	“The principle of Unique Identification: Every entity should be uniquely identified within an identified namespace... The development of domains or namespaces within the Internet has helped in the relaxation of pressure on the need for absolute uniqueness in the structure an identifier. URLs or URIs provide mechanisms for universal disambiguation that allow even common terms to assume unique, global status.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Identifier	INDECS
Orphan works	Works whose copyright holders are unknown and cannot be traced	“works where the copyright owners are not known or cannot be found so it is not possible to get the necessary authorisations to use their works” http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/619	Out of print, In commerce	EU
Out of print	See sample definition	“A publication no longer obtainable through regular market channels because the publisher's inventory is exhausted, with no prospect of another printing in the foreseeable future. A book goes out of print when the publisher decides sales no longer justify the expense of maintaining inventory, when it is superseded by a later edition, or when the rights are relinquished by the publisher.” http://www.abc-clio.com/ODLIS/odlis_o.aspx#outofprint	Content, Edition, In print, Intellectual property, Manifestation, In commerce, Orphan works	LIBRARY SCIENCE
Party	Formal identification of a person or organisation with respect to making or trading intellectual property that is used in Indecs Has a functional parallel in CIDOC's E39 Actor	“An agent undertaking an activity or task in a creative or commercial relation” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf “...people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible.” http://www.cidoc-crm.org/official_release_cidoc.html	Content, Metadata, Intellectual Property, Resource	INDECS

Percept	Entity perceived by the five senses The complement of Indecs' and CIDOC-CRM's top level disjunct, concept(ual object)	"An entity which is perceived directly with at least one of the five senses" http://www.doi.org/topics/indexes/indexes_framework_2000.pdf	Concept	INDECS
Product	In GRid, an Indecs "manifestation", derived from the set of intellectual property more abstractly described in the "release" Otherwise, the set of all identical commercial cultural items	"A Product is a manifestation of a Release, in the form in which it is made available to consumers. The attributes of a Release in its digital manifestation as a Product may be technical (for example, the codec or bit rate); the way in which it is consumed (for example, downloading or streaming); or a commercial term (for example, the price at which it is made available)." http://www.ifpi.org/content/library/GRid_Standard_v2_1.pdf	Manifest, Manifestation, Recording, Release, Resource	GRID ONIX
RDF (Resource Description Framework)	W3C standard data model and syntax recommended for linked (open) data on the Web. It consists of atomic propositional units connected to form more complex or more specific propositions	"RDF is an assertional language intended to be used to express propositions using precise formal vocabularies... for access and use over the World Wide Web, and is intended to provide a basic foundation for more advanced assertional languages with a similar purpose." http://www.w3.org/TR/rdf-mt/#intro "The underlying structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph... The nodes of an RDF graph are its subjects and objects. The assertion of an RDF triple says that some relationship, indicated by the predicate, holds between the things denoted by subject and object of the triple." http://www.w3.org/TR/rdf-concepts/#section-data-model	Conceptual model, Data structure, Identifier, Linked (Open) Data, Metadata, Party	W3C

Record	In cultural heritage contexts, an integral set of metadata elements deemed sufficient for describing a real world item for a given purpose Generally a more or less static database file, and in one format because it originates in one context and is intended for use only in that context	"A record is metadata expressed in a single format." http://www.openarchives.org/OAI/openarchivesprotocol.html#Record "A metadata record consists of a set of attributes, or elements, necessary to describe the resource in question" http://dublincore.org/documents/usageguide/	Choreography, Data element, Schema	OAI DUBLIN CORE
Recording, Sound	A "sound recording" as termed by the ISRC documentation is an infixion of an expression in Index terms	"International Standard Recording Code; ISO standard identifier for audio and video recordings" http://www.doi.org/topics/indexes/indexes_framework_2000.pdf	Expression	ISRC
Reference descriptive metadata	Minimum required metadata for an identifier to be assigned to a product by the registration agency	"In the context of identifiers, a further term of art is required for the metadata that is mandated within an identifier standard. This is sometimes referred to as a "minimum metadata set", but such a description can be misleading (since it simply raises a further question – <i>minimum for what purpose?</i>). We therefore agreed to use the term "reference descriptive metadata" for that set of metadata that has been defined within each standard." http://www.dlib.org/dlib/april06/paskin/04paskin.html	Referent, Registration, Service	DOI

Referent	Entity that is referred to by an identifier –whether perceived or purely conceptual – and specified by a “reference descriptive metadata” record	“A particular object identified by a specific DOI name” http://www.doi.org/handbook_2000/glossary.html	Entity, Identifier, Metadata, Record, Reference descriptive metadata	DOI
Registry / registration (agency)	An actively administered database that provides authority, trust and (at a minimum) resolution services for identifiers.	“The primary role of Registration Agencies (RAs) is to provide services to Registrants – allocating DOI® name prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data.” http://www.doi.org/registration_agencies.html	Identifier, Resolution, Service, Referent, Reference descriptive metadata	DOI
Relation / relator	In (conceptual) data models, an entity (more or less explicitly) that links two or more other entities	“The interaction of percepts and/or concepts; a connection between two or more entities” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Attribute, Data model, Entity, Resource, Referent, Metadata	INDECS

Release	In the DDEX/GRid framework, the package of intellectual property that comprises a “product” – more or less analogous to a table of contents or manifest	“A Release is an abstract entity representing a bundle of one or more Digital Resources compiled by an Issuer for the purpose of electronic distribution to individual consumers, directly or through intermediaries. The Digital Resources in Releases are commonly sound recordings or audio-visual recordings, however, they can also include other Digital Resources (including, for example, text, graphics, software). The Release is not itself a Product. Products have more extensive attributes than Releases; one Release may be disseminated in many different Products.” http://www.ifpi.org/content/section_resources/grid-handbook.html	Abstraction, Manifest, Product, Recording, Resource	GRID
Resolution	Using an identifier to retrieve some kind of information about the thing it identifies. This can be automated or involve some human intervention	“The process in which an identifier is the input (a request) to a network service to receive in return a specific output of one or more pieces of current information (state data) related to the identified entity: e.g. a location (such as URL) where the object can be found. Resolution provides a level of managed indirection between an identifier and the output.” http://www.doi.org/handbook_2000/glossary.html [More information: http://www.doi.org/handbook_2000/resolution.html	Actionability, Identifier, Service	DOI ISBN-A
Resource	The W3C term “resource” (and “entity” as used in Indecs) is universal and explicitly entails only the requirement of identification	“...anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., “today’s weather report for Los Angeles”), as well as a collection of other resources. Not all resources are network “retrievable”; e.g., human beings, corporations, and bound books in a library can also be considered resources...” http://www.w3.org/2003/glossary/keyword/All/resource.html?keywords=resource “We do not limit the scope of what might be a resource. The term “resource” is used in a general sense for whatever might be identified by a URI.” http://www.w3.org/TR/webarch/#id-resources “An entity within the scope of the DOI system; the object may be abstract, physical or digital, as any of these forms may be of relevance in content management (e.g. people, referents, people, agreements)... “any resource” on the Internet” http://www.doi.org/handbook_2000/glossary.html	Entity, Identifier, Metadata, Referent	W3C DOI

Resource map	The full list of the resources grouped together by aggregation – it can mix “real-world” cultural heritage objects with digital representations and other data	“A Resource Map (ReM) is a Resource, with type ore:ResourceMap, with information content consisting of assertions that: MUST describe a single Aggregation. / MUST enumerate the constituent Aggregated Resources. / MAY include additional properties about the Aggregation and Aggregated Resources, such as relationships among the Aggregated Resources, relationships from the Aggregated Resources to other Resources, and other properties.” http://www.openarchives.org/ore/1.0/datamodel#Resource_Map	Aggregation, Cultural heritage object, Release, Resource	OAI-ORE
Schema (metadata schema)	The specification of the allowed elements (with their cardinality), controlled values, and structure of a given type of metadata Defines the class of messages or records in that metadata “language”. Can be expressed in formal ways e.g. in XSD	“The purpose of an XML Schema: Structures schema is to define and describe a class of XML documents by using schema components to constrain and document the meaning, usage and relationships of their constituent parts: datatypes, elements and their content and attributes and their values. Schemas may also provide for the specification of additional document information, such as normalization and defaulting of attribute and element values” http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/structures.html#intro-purpose	Attribute, Controlled vocabulary, Data element, Data structure, Metadata	W3C
Semantics	Definitions of terms with reference to real life use and meaning (“natural language” definitions)	“The primary role of the CRM is to enable information exchange and integration between heterogeneous sources of cultural heritage information...it defines and is restricted to the underlying semantics of database schemata and document structures used in cultural heritage and museum documentation in terms of a formal ontology... it explains the logic of what they actually currently document, and thereby enables semantic interoperability” http://www.cidoc-crm.org/official_release_cidoc.html	Data dictionary	CIDOC-CRM

Service	A pre-determined interaction with a metadata system; at its simplest level it could be the return of a metadata record in response to entering its identifier	“A defined result from a defined action <i>i.e.</i>, do X and the result will be Y” http://www.doi.org/handbook_2000/glossary.html	Actionable (identifier), Registry / registration (agency)	DOI EIDR
Standard	The rules and guidelines produced by trade bodies or by the ISO, to address key characteristics of a resource or process Although originating in different ways, standards all depend on consensus, approval by an independent party, and some process of fact-finding, expert review and final approval	“document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context... Note: Standards should be based on the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits.” http://isotc.iso.org/livelink/livelink?func=ll&objId=4230456	Party, Data model, Metadata schema, Identifier	ISO
Supply chain	The chain of parties, relationships and business processes between the creator and the user of cultural content in a cultural industry	“a supply-chain activity – one which generates an output (a product) based on a managed input (raw materials) which is distributed and transacted with an end-user” http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf	Cultural industries, Party	CDS
Type	The most general attribute of a resource, a basic categorisation or classification	“A categorisation of one or more characteristics of an entity through which it belongs to a group of entities; a characteristic role played by an entity” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Attribute, Medium, Resource	INDECS

Validation	Automated check of conformity of a given metadata record or message to a schema, or identifier to its namespace and/or creation algorithms	“An overriding requirement is that an ONIX for Books Product Information Message must conform to the XML standard, <i>i.e.</i> it must be well-formed XML. It is also a requirement that ONIX messages are valid according to the associated RNG and XSD schemas (which are equivalent).” http://www.editeur.org/files/ONIX%203/Introduction_to_ONIX_for_Books_3.0.pdf	Identifier, Intelligent, Schema, Record	ONIX
Value domain	[see sample definition]	“set of permissible values” http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html	Controlled vocabulary	ISO/IEC 11179-1
Work	Term for a highly abstracted conceptual entity, widely used in FRBR as the top level entity identified In Indecs its use was restricted very narrowly. This report prefers the general term “abstraction” or “creative work” to align with Indecs	“a distinct intellectual or artistic creation... A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work... The concept of what constitutes a work and where the line of demarcation lies between one work and another may in fact be viewed differently from one culture to another.” http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf “in the <indecs> framework a work is a piece of intellectual property (ip) defined directly in terms of the legal provisions of the Berne Convention, so while all works are abstractions, all abstractions are not necessarily works in the legal sense.” http://www.doi.org/topics/indecs/indecs_framework_2000.pdf	Abstraction, Creation, Entity, Intellectual property	INDECS FRBR

Work, Audiovisual	"Audiovisual work" is the term used in ISAN documentation to refer to the Indecs abstractions identified by ISANs.	"International Standard Audiovisual Number; draft ISO standard identifier for audiovisual abstractions" http://www.doi.org/topics/indecs/indecs_framework_2000.pdf "For the purposes of the ISAN, the term "audiovisual work" covers any fixation of motion imagery. The ISAN standard defines "audiovisual work" as follows: audiovisual work: work consisting of a sequence of related images, with or without accompanying sound, which is intended to be made visible as a moving image through the use of devices, regardless of the medium of initial or subsequent fixation." http://www.isan.org/portal/page?_pageid=166,41960&_dad=portal&_schema=PORTAL#FAQ4	Abstraction	ISAN
------------------------------	--	---	-------------	------

13 Appendix 2 – Acronyms used

Acronym	Meaning
AAC	Advanced Audio Coding (digital audio file compression format)
BAPLA	British Association of Picture Libraries and Archives
BnF	Bibliothèque Nationale de France
CEPIC	Coordination of European Picture Agencies Stock, Press and Heritage
CHO	“Cultural Heritage Object” (used of objects or resources contributed to Europeana)
CSV	Comma-Separated Values
CWR	Copyright Works Registration (data format for copyright registration)
DC	Dublin Core (metadata schema and organisation; Dublin Core Metadata Initiative)
DCP	Digital Content Protection
DDEX	Digital Data Exchange
DOI	Digital Object Identifier
EAN	International Article Number (originally European Article Number) AKA barcode n°.
ERN	Electronic Release Notification
EVA	European Visual Artists
Exif	Exchangeable Information File (digital photo file format)
FRBR	Functional Requirements for Bibliographic Records (IFLA standard)
GLAM	Galleries, Libraries, Archives and Museums
GRid	Global Release Identifier
Indecs	Interoperability of Data in E-Commerce Systems
IP	Intellectual Property
IPTC	International Press Telecommunications Council
ISBN	International Standard Book Number
ISO	International Organisation for Standardisation

ISRC	International Standard Recording Code
JPEG	Joint Photographic Experts Group (digital photo file format)
MPEG	Motion Picture Experts Group (audiovisual digital file compression standards)
MP3	MPEG-1 or MPEG-2 Audio Layer III (audio digital file compression standard)
MWG	Metadata Working Group
NISO	National Information Standards Organization
RDF	Resource Description Framework
RRO	Reproduction Rights Organisation
STML	Scientific, technical, medical and legal (publishing)
TIFF	Tagged Image File Format
UPC	Universal Product Code (AKA barcode number).
XMP	eXtensible Metadata Platform

14 Appendix 3 – Sample ONIX records

"The sample message is based on a real book from a real publisher (HarperCollins' 2006 paperback version of *Roseanna* by Maj Sjöwall and Per Wahlöö, ISBN 978-0-00-723283-3)..." (Bell, G., 2011) but this message, its sender and recipient, are fictitious. It is extracted from the ONIX 3.0 guidelines (*ibid.*).

#	ONIX (using reference names/code list codes)	Commentary
1	<?xml version="1.0" encoding="UTF-8"?>	File may contain extended characters without need for special coding
2	<ONIXMessage release="3.0">	
3	<Header>	
4	<Sender>	
5	<SenderName>Global Bookinfo</SenderName>	Fictitious aggregator sending...
6	<ContactName>Jane King, +1 555 321 7654</ContactName>	
7	<EmailAddress>jbk@globalbookinfo.com</EmailAddress>	
8	</Sender>	
9	<Addressee>	
10	<AddresseeName>BooksBooksBooks.com</AddresseeName>	...to fictitious bookseller
11	</Addressee>	
12	<MessageNumber>231</MessageNumber>	Allows recipient to check for missed messages
13	<SentDateTime>20100510T1115-0400</SentDateTime>	10 May 2010, 11:15am EDT (3:15pm UTC)
14	<MessageNote>Sample message</MessageNote>	
15	</Header>	
16	<!-- product record 1 of 1 in message -->	
17	<Product>	Start of product record
18	<RecordReference>com.globalbookinfo.onix.01734529</RecordReference>	Unique record reference assured by using reversed web domain plus internal product ID
19	<NotificationType>03</NotificationType>	Confirmed post-publication
20	<RecordSourceType>04</RecordSourceType>	Bibliographic agency
21	<RecordSourceIdentifier>	
22	<RecordSourceIDType>06</RecordSourceIDType>	GLN
23	<IDValue>0614141800001</IDValue>	
24	</RecordSourceIdentifier>	
25	<RecordSourceName>Global Bookinfo</RecordSourceName>	
26	<ProductIdentifier>	
27	<ProductIDType>03</ProductIDType>	ISBN as GTIN-13
28	<IDValue>9780007232833</IDValue>	
29	</ProductIdentifier>	
30	<ProductIdentifier>	

31	<ProductIDType>15</ProductIDType>	Explicitly an ISBN
32	<IDValue>9780007232833</IDValue>	
33	</ProductIdentifier>	
34	<DescriptiveDetail>	Start of Block 1
35	<ProductComposition>00</ProductComposition>	Single-item retail product
36	<ProductForm>BC</ProductForm>	Paperback
37	<ProductFormDetail>B105</ProductFormDetail>	B-format
38	<Measure>	
39	<MeasureType>01</MeasureType>	Height
40	<Measurement>197</Measurement>	
41	<MeasureUnitCode>mm</MeasureUnitCode>	
42	</Measure>	
43	<Measure>	
44	<MeasureType>02</MeasureType>	Width
45	<Measurement>130</Measurement>	
46	<MeasureUnitCode>mm</MeasureUnitCode>	
47	</Measure>	
48	<Measure>	
49	<MeasureType>03</MeasureType>	Spine thickness
50	<Measurement>18</Measurement>	
51	<MeasureUnitCode>mm</MeasureUnitCode>	
52	</Measure>	
53	<Measure>	
54	<MeasureType>08</MeasureType>	Weight
55	<Measurement>195</Measurement>	
56	<MeasureUnitCode>gr</MeasureUnitCode>	
57	</Measure>	
58	<CountryOfManufacture>GB </CountryOfManufacture>	
59	<ProductClassification>	
60	<ProductClassificationType>03 </ProductClassificationType>	HMRC commodity code
61	<ProductClassificationCode>49019900 </ProductClassificationCode>	
62	</ProductClassification>	
63	<Collection>	Collection detail
64	<CollectionType>10</CollectionType>	Publisher collection
65	<TitleDetail>	
66	<TitleType>01</TitleType>	Distinctive title (of collection)
67	<TitleElement>	
68	<SequenceNumber>2</SequenceNumber>	
69	<TitleElementLevel>02 </TitleElementLevel>	Collection level
70	<TitlePrefix textcase="01">The </TitlePrefix>	
71	<TitleWithoutPrefix textcase="01"> Martin Beck series</TitleWithoutPrefix>	Sentence case
72	</TitleElement>	

73	<TitleElement>	
74	<SequenceNumber>3</SequenceNumber>	
75	<TitleElementLevel>01</TitleElementLevel>	Product level
76	<PartNumber>1</PartNumber>	First in collection
77	</TitleElement>	
78	</TitleDetail>	
79	</Collection>	
80	<TitleDetail>	
81	<TitleType>01</TitleType>	Distinctive title
82	<TitleElement>	
83	<SequenceNumber>1</SequenceNumber>	
84	<TitleElementLevel>01</TitleElementLevel>	Product level
85	<TitleText textcase="01">Roseanna</TitleText>	
86	</TitleElement>	
87	</TitleDetail>	
88	<Contributor>	
89	<SequenceNumber>1</SequenceNumber>	Contributor 1
90	<ContributorRole>A01</ContributorRole>	Written by
91	<NameIdentifier>	
92	<NameIDType>01</NameIDType>	Proprietary
93	<IDTypeName>HCP Author ID</IDTypeName>	
94	<IDValue>7421</IDValue>	
95	</NameIdentifier>	
96	<NamesBeforeKey>Maj</NamesBeforeKey>	
97	<KeyNames>Sjöwall</KeyNames>	
98	<BiographicalNote textformat="05"> Maj Sjöwall was born in Stockholm in 1935. She is a poet, novelist and translator, and is best known for the ten Martin Beck novels she wrote with husband Per Wahlöö. </BiographicalNote>	
99	</Contributor>	
100	<Contributor>	
101	<SequenceNumber>2</SequenceNumber>	Contributor 2
102	<ContributorRole>A01</ContributorRole>	Written by
103	<NameIdentifier>	
104	<NameIDType>01</NameIDType>	Proprietary
105	<IDTypeName>HCP Author ID</IDTypeName>	
106	<IDValue>7422</IDValue>	
107	</NameIdentifier>	
108	<NamesBeforeKey>Per</NamesBeforeKey>	
109	<KeyNames>Wahlöö</KeyNames>	
110	<BiographicalNote textformat="05"><p>Per Wahlöö was born in Göteborg. After graduating from the University of Lund in 1946, he worked as a	

journalist, covering criminal and social issues for a number of newspapers and magazines. In the 1950s, Wahlöö became involved with radical political causes, activities that resulted in his deportation from Franco's Spain in 1957. After returning to Sweden, he wrote a number of television and radio plays, and was managing editor of several magazines, before becoming a full-time writer. He is best known for the series of ten <i>Martin Beck</i> novels he wrote with wife Maj Sjöwall, which they completed immediately before his death in 1975.		
111	</Contributor>	
112	<Contributor>	
113	<SequenceNumber>3</SequenceNumber>	Contributor 3
114	<ContributorRole>B06</ContributorRole>	Translator
115	<FromLanguage>swe</FromLanguage>	from Swedish
116	<NameIdentifier>	
117	<NameIDType>01</NameIDType>	Proprietary
118	<IDTypeName>HCP Author ID</IDTypeName>	
119	<IDValue>11150</IDValue>	
120	</NameIdentifier>	
121	<NamesBeforeKey>Lois</NamesBeforeKey>	
122	<KeyNames>Roth</KeyNames>	
123	</Contributor>	
124	<Contributor>	
125	<SequenceNumber>4</SequenceNumber>	Contributor 4
126	<ContributorRole>A24</ContributorRole>	Introduction by
127	<NameIdentifier>	
128	<NameIDType>01</NameIDType>	Proprietary
129	<IDTypeName>HCP Author ID</IDTypeName>	
130	<IDValue>11151</IDValue>	
131	</NameIdentifier>	
132	<NamesBeforeKey>Henning</NamesBeforeKey>	
133	<KeyNames>Mankell</KeyNames>	
134	</Contributor>	
135	<ContributorStatement>By Maj Sjöwall and Per Wahlöö, translated by Lois Roth and with an introduction by Henning Mankell</ContributorStatement>	Text may used for display by websites
136	<NoEdition/>	
137	<Language>	No default language in header
138	<LanguageRole>01</LanguageRole>	Language of text
139	<LanguageCode>eng</LanguageCode>	English
140	</Language>	
141	<Language>	
142	<LanguageRole>02</LanguageRole>	Language of original
143	<LanguageCode>swe</LanguageCode>	Swedish

144	</Language>	
145	<Extent>	
146	<ExtentType>00</ExtentType>	Main content page count
147	<ExtentValue>245</ExtentValue>	
148	<ExtentUnit>03</ExtentUnit>	Pages
149	</Extent>	
150	<Extent>	
151	<ExtentType>04</ExtentType>	Back matter page count
152	<ExtentValue>16</ExtentValue>	
153	<ExtentUnit>03</ExtentUnit>	Pages
154	</Extent>	
155	<Subject>	
156	<MainSubject/>	
157	<SubjectSchemeIdentifier>12</SubjectSchemeIdentifier>	BIC subject category code
158	<SubjectSchemeVersion>2.0</SubjectSchemeVersion>	Version 2
159	<SubjectCode>FF</SubjectCode>	Crime and mystery
160	</Subject>	
161	<Subject>	
162	<SubjectSchemeIdentifier>12</SubjectSchemeIdentifier>	BIC subject category code
163	<SubjectSchemeVersion>2.0</SubjectSchemeVersion>	
164	<SubjectCode>FYT</SubjectCode>	Fiction in translation
165	</Subject>	
166	<Subject>	
167	<SubjectSchemeIdentifier>13</SubjectSchemeIdentifier>	BIC geographical qualifier
168	<SubjectSchemeVersion>2.0</SubjectSchemeVersion>	
169	<SubjectCode>1DNS</SubjectCode>	Sweden
170	</Subject>	
171	<Subject>	
172	<SubjectSchemeIdentifier>15</SubjectSchemeIdentifier>	BIC time period qualifier
173	<SubjectSchemeVersion>2.0</SubjectSchemeVersion>	
174	<SubjectCode>3JJPK</SubjectCode>	1960s
175	</Subject>	
176	<Subject>	
177	<MainSubject/>	
178	<SubjectSchemeIdentifier>10</SubjectSchemeIdentifier>	BISAC subject heading code
179	<SubjectSchemeVersion>2009</SubjectSchemeVersion>	2009 version
180	<SubjectCode>FIC022000</SubjectCode>	Mystery and detective
181	</Subject>	
182	<Audience>	

183	<AudienceCodeType>01</AudienceCodeType>	ONIX audience code
184	<AudienceCodeValue>01</AudienceCodeValue>	General/trade
185	</Audience>	
186	</DescriptiveDetail>	End of Block 1
187	<CollateralDetail>	Start of Block 2
188	<TextContent>	
189	<TextType>02</TextType>	Short description
190	<ContentAudience>00</ContentAudience>	Any audience
191	<Text textformat="05"> Perennial relaunches the first novel in the classic Martin Beck detective series from the 1960s - the novels that have inspired all crime fiction written ever since. </Text>	XHTML markup
192	</TextContent>	
193	<TextContent>	
194	<TextType>03</TextType>	(Long) Description
195	<ContentAudience>00</ContentAudience>	Any audience
196	<Text textformat="05"> Widely recognised as the among the greatest crime fiction ever written, this is the first of a series of stories that pioneered the police procedural genre. The series was translated into 35 languages, sold over 10 million copies around the world, and inspired writers from Henning Mankell to Jonathan Franzen. Written in 1965, Roseanna is the work of Maj Sjöwall and Per Wahlöö - a husband and wife team from Sweden, and this volume has a new introduction to help bring their work to a new audience. The novel follows the fortunes of the detective Martin Beck, whose enigmatic and taciturn character has inspired countless other policemen in crime fiction. Roseanna begins on a July afternoon: the body of a young woman is dredged from a canal near Sweden's beautiful Lake Vättern. Three months later, all that Police Inspector Martin Beck knows is that her name is Roseanna, that she came from Lincoln, Nebraska, and that she could have been strangled by any one of eighty- five people. With its authentically rendered settings and vividly realized characters, and its command over the intricately woven details of police detection, Roseanna is a masterpiece of suspense and sadness. </Text>	Multiple paragraphs
197	</TextContent>	
198	<TextContent>	
199	<TextType>06</TextType>	Review quote
200	<ContentAudience>00</ContentAudience>	Any audience
201	<Text textformat="05"> 'The writing is elegant and surprisingly humorous - if you haven't come across Beck before, you're in for a treat.' </Text>	
202	<SourceTitle>The Guardian</SourceTitle>	

203	</TextContent>	
204	<TextContent>	
205	<TextType>06</TextType>	Review quote
206	<ContentAudience>00</ContentAudience>	Any audience
207	<Text textformat="05"> 'Their mysteries don't just read well; they reread even better. Witness, wife, petty cop or crook – they're all real characters even if they get just a few sentences. The plots hold, because they're ingenious but never inhuman.' </Text>	
208	<SourceTitle>New York Times</SourceTitle>	
209	</TextContent>	
210	<CitedContent>	
211	<CitedContentType>03</CitedContentType>	Media mention (feature article)
212	<ContentAudience>00</ContentAudience>	Any audience
213	<SourceType>01</SourceType>	Printed media
214	<SourceTitle>Observer Magazine</SourceTitle>	
215	<CitationNote>Interview with Maj Sjöwall by Louise France</CitationNote>	
216	<ResourceLink> http://www.guardian.co.uk/books/2009/nov/22/crime-thriller-maj-sjowall-sweden </ResourceLink>	URL
217	<ContentDate>	
218	<ContentDateRole>01</ContentDateRole>	Publication date (of article)
219	<Date dateformat="00">20091122</Date>	
220	</ContentDate>	
221	</CitedContent>	
222	<SupportingResource>	
223	<ResourceContentType>01</ResourceContentType>	Front cover
224	<ContentAudience>00</ContentAudience>	All audiences
225	<ResourceMode>03</ResourceMode>	Image
226	<ResourceVersion>	
227	<ResourceForm>02</ResourceForm>	Downloadable
228	<ResourceVersionFeature>	
229	<ResourceVersionFeatureType>01</ResourceVersionFeatureType>	File format
230	<FeatureValue>D502</FeatureValue>	JPEG
231	</ResourceVersionFeature>	
232	<ResourceVersionFeature>	
233	<ResourceVersionFeatureType>02</ResourceVersionFeatureType>	Image height
234	<FeatureValue>341</FeatureValue>	in pixels
235	</ResourceVersionFeature>	
236	<ResourceVersionFeature>	
237	<ResourceVersionFeatureType>03</ResourceVersionFeatureType>	Image width
238	<FeatureValue>222</FeatureValue>	in pixels

239	</ResourceVersionFeature>	
240	<ResourceLink> http://www.harpercollins.co.uk/covers/9780007232833.jpg </ResourceLink>	URL
241	<ContentDate>	
242	<ContentDateRole>17</ContentDateRole>	Last updated
243	<Date dateformat="00">20060412</Date>	
244	</ContentDate>	
245	</ResourceVersion>	
246	</SupportingResource>	
247	<SupportingResource>	
248	<ResourceContentType>15</ResourceContentType>	Sample content
249	<ContentAudience>00</ContentAudience>	All audiences
250	<ResourceMode>06</ResourceMode>	Multi-mode (a web page)
251	<ResourceVersion>	
252	<ResourceForm>01</ResourceForm>	Linkable resource
253	<ResourceLink> http://browseinside.harpercollins.co.uk/pageview?isbn=9780007232833 </ResourceLink>	URL
254	</ResourceVersion>	
255	</SupportingResource>	
256	</CollateralDetail>	End of Block 2
257	<!-- there is no Block 3 -->	
258	<PublishingDetail>	Start of Block 4
259	<Imprint>	
260	<ImprintName>HarperPerennial</ImprintName>	
261	</Imprint>	
262	<Publisher>	
263	<PublishingRole>01</PublishingRole>	Publisher
264	<PublisherName>HarperCollins Publishers</PublisherName>	
265	<Website>	
266	<WebsiteRole>01</WebsiteRole>	Corporate site
267	<WebsiteLink> http://www.harpercollins.co.uk </WebsiteLink>	
268	</Website>	
269	</Publisher>	
270	<CityOfPublication>London</CityOfPublication>	
271	<CountryOfPublication>GB</CountryOfPublication>	
272	<PublishingStatus>04</PublishingStatus>	Active
273	<PublishingDate>	
274	<PublishingDateRole>01</PublishingDateRole>	Date of publication (of product)
275	<Date dateformat="00">20060807</Date>	dateformat=YYYYMMDD
276	</PublishingDate>	
277	<PublishingDate>	
278	<PublishingDateRole>11</PublishingDateRole>	Date of first publication (of

	</PublishingDateRole>	work)
279	<Date dateformat="05">1968</Date>	dateformat=YYYY
280	</PublishingDate>	
281	<PublishingDate>	
282	<PublishingDateRole>20 </PublishingDateRole>	Pub date in original language
283	<Date dateformat="05">1965</Date>	dateformat=YYYY
284	</PublishingDate>	
285	<SalesRights>	
286	<SalesRightsType>01</SalesRightsType>	For sale (publisher has exclusive rights in territory)
287	<Territory>	
288	<CountriesIncluded>AG AI AU BB BD BM BN BS BW BZ CM CY DM EG FJ FK GB GD GG GH GI GM GY HK IE IL IM IN IO JE JM JO KE KI KN KW KY LC LK LS MT MU MV MW MY MZ NA NG NR NZ PG PK PN SB SC SD SG SH SL SO SZ TC TO TT TV TZ UG VC VG VU WS YE RS ME ZA ZM ZW</CountriesIncluded>	Broadly, this is 'Commonwealth excluding Canada'
289	</Territory>	
290	</SalesRights>	
291	<SalesRights>	
292	<SalesRightsType>06</SalesRightsType>	Not for sale (publisher has no rights in territory)
293	<Territory>	
294	<CountriesIncluded>AS CA GU MP PH PR US VI</CountriesIncluded>	Broadly, the 'US Market' and Canada
295	</Territory>	
296	</SalesRights>	
297	<ROWSalesRightsType>02</ROWSalesRightsType>	For sale (publisher has non-exclusive rights in rest of world)
298	</PublishingDetail>	End of Block 4
299	<RelatedMaterial>	Start of Block 5
300	<RelatedWork>	
301	<WorkRelationCode>01</WorkRelationCode>	Is manifestation of
302	<WorkIdentifier>	
303	<WorkIDType>11</WorkIDType>	ISTC
304	<IDValue>A0220090000154FA</IDValue>	
305	</WorkIdentifier>	
306	</RelatedWork>	
307	<RelatedProduct>	
308	<ProductRelationCode>06 </ProductRelationCode>	Alternative format
309	<ProductRelationCode>27 </ProductRelationCode>	E-version
310	<ProductIdentifier>	
311	<ProductIDType>03</ProductIDType>	GTIN-13
312	<IDValue>9780007324378</IDValue>	
313	</ProductIdentifier>	

314	<ProductIdentifier>	
315	<ProductIDType>15</ProductIDType>	ISBN
316	<IDValue>9780007324378</IDValue>	
317	</ProductIdentifier>	
318	</RelatedProduct>	
319	</RelatedMaterial>	End of Block 5
320	<ProductSupply>	Start of Block 6
321	<Market>	
322	<Territory>	
323	<RegionsIncluded>WORLD</RegionsIncluded>	Market is world...
324	<CountriesExcluded>AS AU CA GU MP NZ PH PR US VI ZA</CountriesExcluded>	...minus countries with no rights and countries with separate exclusive distributorships (the latter could additionally be represented by separate <ProductSupply> sections)
325	</Territory>	
326	</Market>	
327	<MarketPublishingDetail>	
328	<MarketPublishingStatus>04</MarketPublishingStatus>	Active
329	<MarketDate>	
330	<MarketDateRole>01</MarketDateRole>	Pub date in this market
331	<Date dateformat="00">20060807</Date>	dateformat=YYYYMMDD
332	</MarketDate>	
333	</MarketPublishingDetail>	
334	<SupplyDetail>	
335	<Supplier>	
336	<SupplierRole>01</SupplierRole>	Publisher supplying to retailers
337	<SupplierIdentifier>	
338	<SupplierIDType>06</SupplierIDType>	GLN for orders
339	<IDValue>5051366000000</IDValue>	
340	</SupplierIdentifier>	
341	<SupplierIdentifier>	
342	<SupplierIDType>07</SupplierIDType>	SAN for orders
343	<IDValue>0091073</IDValue>	
344	</SupplierIdentifier>	
345	<SupplierName>HarperCollins Publishers</SupplierName>	
346	<TelephoneNumber>+44 1417 723200</TelephoneNumber>	
347	</Supplier>	
348	<ReturnsConditions>	
349	<ReturnsCodeType>02</ReturnsCodeType>	BISAC indicator
350	<ReturnsCode>Y</ReturnsCode>	Returnable
351	</ReturnsConditions>	
352	<ProductAvailability>21</ProductAvailability>	In stock

	</ProductAvailability>	
353	<PackQuantity>16</PackQuantity>	16 copies in a carton
354	<Price>	Price 1 for UK
355	<PriceType>02</PriceType>	RRP including tax
356	<DiscountCoded>	
357	<DiscountCodeType>01</DiscountCodeType>	BIC discount group code
358	<DiscountCode>AHACP029</DiscountCode>	HCP discount group 29
359	</DiscountCoded>	
360	<PriceStatus>02</PriceStatus>	Firm price
361	<PriceAmount>7.99</PriceAmount>	
362	<Tax>	
363	<TaxType>01</TaxType>	VAT
364	<TaxRateCode>Z</TaxRateCode>	Zero-rated
365	<TaxRatePercent>0</TaxRatePercent>	0%
366	<TaxableAmount>7.99</TaxableAmount>	
367	<TaxAmount>0.00</TaxAmount>	
368	</Tax>	
369	<CurrencyCode>GBP</CurrencyCode>	Pounds Sterling
370	<Territory>	Price applicable to...
371	<CountriesIncluded>GB</CountriesIncluded>	
372	</Territory>	
373	<PrintedOnProduct>02</PrintedOnProduct>	Price on product
374	<PositionOnProduct>01</PositionOnProduct>	On outside back cover
375	</Price>	
376	<Price>	Price 2 for Eurozone
377	<PriceType>01</PriceType>	RRP excluding tax
378	<Discount>	
379	<DiscountPercent>37.5</DiscountPercent>	37.5% discount
380	</Discount>	
381	<PriceStatus>02</PriceStatus>	Firm price
382	<PriceAmount>8.99</PriceAmount>	
383	<CurrencyCode>EUR</CurrencyCode>	Euro
384	<Territory>	Price applicable to...
385	<CountriesIncluded>AT BE CY FI FR DE ES GR IE IT LU MT NL PT SI SK AD MC ME SM VA</CountriesIncluded>	Euro-using countries in Europe
386	</Territory>	
387	<PrintedOnProduct>01</PrintedOnProduct>	Price not on product
388	</Price>	
389	<Price>	Price 3 for remainder of market
390	<PriceType>01</PriceType>	RRP excluding tax
391	<Discount>	
392	<DiscountPercent>42.5</DiscountPercent>	42.5% discount

	</DiscountPercent>	
393	</Discount>	
394	<PriceStatus>02</PriceStatus>	Firm price
395	<PriceAmount>7.99</PriceAmount>	
396	<CurrencyCode>GBP</CurrencyCode>	Pounds Sterling
397	<Territory>	Price applicable to...
398	<RegionsIncluded>WORLD </RegionsIncluded>	Remainder of market, ie world...
399	<CountriesExcluded>GB AT BE CY FI FR DE ES GR IE IT LU MT NL PT SI SK AD MC ME SM VA AS AU CA GU MP NZ PH PR US VI ZA</CountriesExcluded>	...minus GB, Euro-using countries in Europe and countries not in market
400	</Territory>	
401	<PrintedOnProduct>01</PrintedOnProduct>	Price not on product
402	</Price>	
403	</SupplyDetail>	
404	</ProductSupply>	End of Block 6
405	<!-- aggregator could add other ProductSupply sections here, for AU/NZ and ZA distributors -->	
406	</Product>	
407	</ONIXMessage>	

15 Appendix 4 – Comparison of photo metadata formats

IPTC Core 1.1	IPTC Extension 1.1	IPTC-NAA IIM 4.1	XMP namespace:property ID	FotoStation 7.0	Photoshop CS3	Photo Mechanic 4.6.2	Definition
Generic Name	Generic Name	Application Record Number and Name		v4.5 Style Editor	[name of pane]		
Creator		2:80 By-line	dc:creator	Byline	Author [Description]	Photographer	Byline
Description		2:120 Caption/Abstract	dc:description	Caption	Description [Description]	Caption	Caption
Copyright Notice		2:116 Copyright Notice	dc:rights	Copyright String	Copyright Notice [Description]	Copyright	Copyright
Keywords		2:25 Keywords	dc:subject	Keywords	Keywords [Description]	Keywords	Keywords
Title		2:05 Object Name	dc:title	Object Name	Document Title [Description]	Object Name	Normally used to store the filename of the image; not the same as Headline.
Country Code (legacy)		2:100 Country/Primary Location Code	iptc4xmpCore:CountryCode	Country Code	ISO Country Code [IPTC Image]	Code	Country ISO Code (ISO 3166-1 alpha-3)
Sublocation (legacy)		2:92 Sublocation	iptc4xmpCore:Location	n/a	Location [IPTC Image]	Location	Specific location, such as Madison Square Garden, Kodak Theatre, Grand Canyon etc...

IPTC Core 1.1	IPTC Extension 1.1	IPTC-NAA IIM 4.1	XMP namespace:property ID	FotoStation 7.0	Photoshop CS3	Photo Mechanic 4.6.2	Definition
Creator's jobtitle		2:85	By-line Title	photoshop:AuthorsPosition	Byline Title	Author Title [Description]	Byline Title
Description Writer		2:122	Writer/Editor	photoshop:CaptionWriter	Caption Writer	Description Writer [Description]	Initials of Caption Writer
		2:15	Category <marked as deprecated>	photoshop:Category	Category	Category [Categories]	Category
City (legacy)		2:90	City	photoshop:City	City	City [Origin]	City
Country (legacy)		2:101	Country/Primary Location Name	photoshop:Country	Country	Country [Origin]	Country Name
Credit Line		2:110	Credit	photoshop:Credit	Credit	Credit [Origin]	Credit
Date Created		2:55	Date Created	photoshop:DateCreated	IPTC Created Date	Date Created [Origin]	Create Date
Headline		2:105	Headline	photoshop:Headline	Headline	Headline [Origin]	Headline / Event Name
Instructions		2:40	Special Instructions	photoshop:Instructions	Special Instructions	Instructions [Origin]	Special Instructions
Source		2:115	Source	photoshop:Source	Source	Source [Origin]	Source
Province or State (legacy)		2:95	Province/State	photoshop:State	Province State	State/Province [Origin]	State or Province
		2:20	Supplemental Category <marked as deprecated>	photoshop:SupplementalCategories	Supplemental Categories	Supplemental Categories [Categories]	Supplemental Categories
Job Id		2:103	Original Transmission Reference	photoshop:TransmissionReference	Original Transmission Reference	Transmission Reference [Origin]	Media Event ID (7-9 numeric chrs)

IPTC Core 1.1	IPTC Extension 1.1	IPTC-NAA IIM 4.1	XMP namespace:property ID	FotoStation 7.0	Photoshop CS3	Photo Mechanic 4.6.2	Definition	
		<defined as field number 121 in FotoStation>		Local Caption	<delimited inside of Description>	<delimited inside of Caption>	(Delimiter: *** Local Caption ***)	
		2:22	Fixture Identifier	Fixture Identifier				
		2:65	Originating Program	Originating Progr...				
		2:10	Urgency <marked as deprecated>	photoshop:Urgency	Priority	Urgency [Origin]	Urgency	Editorial urgency or transmission priority
	Additional Model Information		lptc4xmpExt:AddlModelInfo					
	Artwork or Object in the Image		lptc4xmpExt:ArtworkOrObject					
	Code of Organisation Featured in the Image		lptc4xmpExt:OrganisationInImageCode					
	Controlled Vocabulary Term		lptc4xmpExt:CVterm					
	Location Shown in the Image		lptc4xmpExt:LocationShown					
	Model Age		lptc4xmpExt:ModelAge					

IPTC Core 1.1	IPTC Extension 1.1	IPTC-NAA IIM 4.1	XMP namespace:property ID	FotoStation 7.0	Photoshop CS3	Photo Mechanic 4.6.2	Definition
	Name of Organisation Featured in the Image		Iptc4xmpExt:OrganisationInImageName				
	Person Shown in the Image		Iptc4xmpExt:PersonInImage				
	Digital Image GUID		Iptc4xmpExt:DigImageGUID				
	Digital Source Type		Iptc4xmpExt:DigitalSourceType				
	Event		Iptc4xmpExt:Event				
	Image Registry Entry		Iptc4xmpExt:RegistryId				
	Image Supplier		Iptc4xmpExt:ImageSupplier				
	Location created		Iptc4xmpExt:LocationCreated				
	Max Avail Height		Iptc4xmpExt:MaxAvailHeight				
	Max Avail Width		Iptc4xmpExt:MaxAvailWidth				
	PLUS Version		plus:Version				
	Copyright Owner		plus:CopyrightOwner				
	Image Creator		plus:ImageCreator				
	Licensor		plus:Licensor				

IPTC Core 1.1	IPTC Extension 1.1	IPTC-NAA IIM 4.1	XMP namespace:property ID	FotoStation 7.0	Photoshop CS3	Photo Mechanic 4.6.2	Definition
	Minor Model Age Disclosure		plus:MinorModelAgeDisclosure				
	Model Release Id		plus:ModelReleaseID				
	Model Release Status		plus:ModelReleaseStatus				
	Property Release Id		plus:PropertyReleaseID				
	Property Release Status		plus:PropertyReleaseStatus				
* http://iptc.org/		* http://iptc.org/		* http://fotoware.com	* http://adobe.com		
last update: 07/22/2011							

Table contributed by: Andreas Gebhard

Twitter: @agebhard Web: <http://andreasgebhard.com/>

16 Appendix 5 – Sample Exif, IPTC and XMP records

Example from picture library (from <https://libraries.ucsd.edu/apps/public/#ark:bb12297050> and courtesy of Greg Reser - greser@ucsd.edu) showing possible uses of all fields.

---- File ----	
FileName	EMwg-IPTC_all.jpg
Directory	.
FileSize	370 kB
FileModifyDate	2011:11:08 16:43:51+00:00
FilePermissions	rw-rw-rw-
FileType	JPEG
MIMEType	image/jpeg
ExifByteOrder	Big-endian (Motorola, MM)
CurrentIPTCDigest	a0417e45744a32eeb068976cc607dc4f
ImageWidth	768
ImageHeight	762
EncodingProcess	Baseline DCT, Huffman coding
BitsPerSample	8
ColorComponents	1
---- JFIF ----	
JFIFVersion	1.01
ResolutionUnit	inches
XResolution	1500
YResolution	1500
---- EXIF ----	
ImageDescription	Central University Library; Digital Object URL and Metadata: https://libraries.ucsd.edu/ark:/20775/bb12297050

Orientation	Horizontal (normal)
Artist	SIO Photographic Laboratory
Copyright	Copyright UC Regents; This work is protected by the U.S. Copyright Law (Title 17, U.S.C.). Use of this work beyond that allowed by "fair use" requires written permission of the UC Regents. Permission may be obtained from the UC San Diego Libraries department having custody of the work (http://libraries.ucsd.edu/locations/sio/). Responsibility for obtaining permissions and any use and distribution of this work rests exclusively with the user and not the UC San Diego Libraries.
ExifVersion	0221
ColorSpace	Uncalibrated
---- XMP ----	
XMPToolkit	Adobe XMP Core 4.2-c020 1.124078, Tue Sep 11 2007 23:21:40
DocumentID	4F1DECB41A60DC5A37954A90C061B8E0
OriginalDocumentID	4F1DECB41A60DC5A37954A90C061B8E0
InstanceID	xmp.iid:61F303AD10F1E0118B00A540531ACE9B
Format	image/jpeg
LegacyIPTCDigest	A0417E45744A32EEB068976CC607DC4F
AuthorsPosition	Photographer
DateCreated	1970:12:10
City	La Jolla
State	California
Country	United States
Headline	UCSD Giesel Library as it appeared in 1970 when it was called "Central Library"
CaptionWriter	Greg Reser
TransmissionReference	order 6985-1

Instructions	This work may be used under U.S. "Fair Use" restrictions
Credit	Scripps Institution of Oceanography Library, University of California, San Diego, La Jolla, CA 92093-0219 (http://libraries.ucsd.edu/locations/sio/)
Source	Scripps Institution of Oceanography Library Archives
ColorMode	Grayscale
ICCProfileName	
IntellectualGenre	genre:Archive_material
Location	University of California, San Diego
CountryCode	US
MetadataDate	2011:11:08 08:43:51-08:00
ExifVersion	0221
ExifImageWidth	768
ExifImageHeight	762
ColorSpace	Uncalibrated
Orientation	Horizontal (normal)
ImageWidth	768
ImageHeight	762
PhotometricInterpretation	BlackIsZero
SamplesPerPixel	4
XResolution	72
YResolution	72
ResolutionUnit	inches
AlreadyApplied	True
PropertyReleaseStatus	None

PLUSVersion	1.2.0
HistoryAction	saved
HistoryInstanceID	xmp.iid:61F303AD10F1E0118B00A540531ACE9B
HistoryWhen	2011:10:07 11:20:51-07:00
HistorySoftwareAgent	Adobe Photoshop Camera Raw 6.4.1
HistoryChanged	/metadata
Description	Central University Library; Digital Object URL and Metadata: https://libraries.ucsd.edu/ark:/20775/bb12297050
Creator	SIO Photographic Laboratory
Subject	University of California, San Diego, Archicet cure, Buildings, Libraries, Higher education institutions, Campuses, Archtectural photography
Title	UCSD Central University Library, October 28, 1970
Rights	© Copyright UC Regents; This work is protected by the U.S. Copyright Law (Title 17, U.S.C.). Use of this work beyond that allowed by "fair use" requires written permission of the UC Regents. Permission may be obtained from the UC San Diego Libraries department having custody of the work (http://libraries.ucsd.edu/locations/sio/). Responsibility for obtaining permissions and any use and distribution of this work rests exclusively with the user and not the UC San Diego Libraries.
CreatorAddress	9500 Gilman Drive
CreatorCity	La Jolla
CreatorRegion	California
CreatorPostalCode	92093
CreatorCountry	United States
CreatorWorkTelephone	(858) 534-3274
CreatorWorkEmail	silib@sio.ucsd.edu
CreatorWorkURL	http://www.libraries.ucsd.edu/locations/sio/

Scene	011000, 011600, 01140
SubjectCode	01002000, 01009000, 04018001, 05000000, 0500700
BitsPerSample	8, 8, 8, 8
LicensorName	Scripps Institution of Oceanography Library, University of California, San Diego
LicensorID	96039850
LicensorStreetAddress	9500 Gilman Drive
LicensorExtendedAddress	#0219
LicensorCity	La Jolla
LicensorRegion	North America
LicensorPostalCode	92037-0219
LicensorCountry	United States
LicensorTelephone1	(858) 534-3274
LicensorTelephone2	(858) 822-3074
LicensorEmail	siolib@sio.ucsd.edu
LicensorURL	http://libraries.ucsd.edu/locations/sio/
CopyrightOwnerName	University of California Regents
CopyrightOwnerID	85088322
ImageCreatorID	96039850-1
LocationCreatedSublocation	University of California, San Diego
LocationCreatedCity	La Jolla
LocationCreatedProvinceState	California
LocationCreatedCountryName	United States
LocationCreatedCountryCode	US
LocationCreatedWorldRegion	North America
LocationShownSublocation	University of California, San Diego

LocationShownCity	La Jolla
LocationShownProvinceState	California
LocationShownCountryName	United States
LocationShownCountryCode	US
LocationShownWorldRegion	North America
Event	UCSD Central Library Dedication
OrganisationInImageName	University of California, San Diego
ArtworkDateCreated	1970
ArtworkCopyrightNotice	Public Domain
ArtworkTitle	University Library Building
ArtworkCreator	William L. Pereira Associates
UsageTerms	This work is available from the UC San Diego Libraries. This digital copy of the work is intended to support research, teaching, and private study.
---- IPTC ----	
ApplicationRecordVersion	2
ObjectName	UCSD Central University Library, October 28, 1970
Keywords	University of California, San Diego, Archiceture, Buildings, Libraries, Higher education institutions, Campuses, Archtectural photography
SpecialInstructions	This work may be used under U.S. "Fair Use" restrictions
DateCreated	1970:12:10
By-line	SIO Photographic Laboratory
By-lineTitle	Photographer
City	La Jolla
Sub-location	University of California, San Di
Province-State	California

Country-PrimaryLocationName	United States
OriginalTransmissionReference	order 6985-1
Headline	UCSD Giesel Library as it appeared in 1970 when it was called "Central Library"
Credit	Scripps Institution of Oceanogra
Source	Scripps Institution of Oceanogra
CopyrightNotice	© Copyright UC Regents; This work is protected by the U.S. Copyright Law (Title 17, U.S.C.). Use of this work beyond that allowe
Caption-Abstract	Central University Library; Digital Object URL and Metadata: https://libraries.ucsd.edu/ark:/20775/bb12297050
Writer-Editor	Greg Reser
---- Photoshop ----	
IPTCDigest	a0417e45744a32eeb068976cc607dc4f

Example of very comprehensive (but not using absolutely all fields) photo data from a picture agency perspective (see <http://www.iptc.org/std/photometadata/documentation/GenericGuidelines/> - and courtesy of David Riecks - david@riecks.com).

---- File ----	
FileName	staff-photographer-example.jpg
Directory	.
FileSize	107 kB
FileModifyDate	2010:07:16 03:24:34+01:00
FilePermissions	rw-rw-rw-
FileType	JPEG
MIMETYPE	image/jpeg
ExifByteOrder	Little-endian (Intel, II)
CurrentIPTCDigest	0af9963abe0645b9bba74dcf1832ac7c

ImageWidth	516
ImageHeight	350
EncodingProcess	Baseline DCT, Huffman coding
BitsPerSample	8
ColorComponents	3
YCbCrSubSampling	YCbCr4:4:4 (1 1)
---- EXIF ----	
ImageWidth	806
ImageHeight	547
BitsPerSample	16 16 16
Compression	Uncompressed
PhotometricInterpretation	RGB
ImageDescription	After digging the furrows another ten yards with the tractor, Jim Moore hops off to hand-set more leeks and onions.
Orientation	Horizontal (normal)
SamplesPerPixel	3
XResolution	72
YResolution	72
PlanarConfiguration	Chunky
ResolutionUnit	inches
Software	Adobe Photoshop CS5 Windows
ModifyDate	2010:07:14 03:38:26
Artist	John Doe
Copyright	©2010 Big Newspaper, all rights reserved
ExifVersion	0221
ColorSpace	sRGB

ExifImageWidth	516
ExifImageHeight	350
Compression	JPEG (old-style)
XResolution	72
YResolution	72
ResolutionUnit	inches
ThumbnailOffset	618
ThumbnailLength	5623
---- ICC_Profile ----	
ProfileCMType	Lino
ProfileVersion	2.1.0
ProfileClass	Display Device Profile
ColorSpaceData	RGB
ProfileConnectionSpace	XYZ
ProfileDateTime	1998:02:09 06:49:00
ProfileFileSignature	acsp
PrimaryPlatform	Microsoft Corporation
CMMFlags	Not Embedded, Independent
DeviceManufacturer	IEC
DeviceModel	sRGB
DeviceAttributes	Reflective, Glossy, Positive, Color
RenderingIntent	Media-Relative Colorimetric
ConnectionSpaceIlluminant	0.9642 1 0.82491
ProfileCreator	HP
ProfileID	0
ProfileCopyright	Copyright (c) 1998 Hewlett-Packard Company

ProfileDescription	sRGB IEC61966-2.1
MediaWhitePoint	0.95045 1 1.08905
MediaBlackPoint	0 0 0
RedMatrixColumn	0.43607 0.22249 0.01392
GreenMatrixColumn	0.38515 0.71687 0.09708
BlueMatrixColumn	0.14307 0.06061 0.7141
DeviceMfgDesc	IEC http://www.iec.ch
DeviceModelDesc	IEC 61966-2.1 Default RGB colour space – sRGB
ViewingCondDesc	Reference Viewing Condition in IEC61966-2.1
ViewingCondIlluminant	19.6445 20.3718 16.8089
ViewingCondSurround	3.92889 4.07439 3.36179
ViewingCondIlluminantType	D50
Luminance	76.03647 80 87.12462
MeasurementObserver	CIE 1931
MeasurementBacking	0 0 0
MeasurementGeometry	Unknown (0)
MeasurementFlare	0.999%
MeasurementIlluminant	D65
Technology	Cathode Ray Tube Display
RedTRC	(Binary data 2060 bytes, use -b option to extract)
GreenTRC	(Binary data 2060 bytes, use -b option to extract)
BlueTRC	(Binary data 2060 bytes, use -b option to extract)
---- APP14 ----	
DCTEncodeVersion	100
APP14Flags0	[14]
APP14Flags1	(none)

ColorTransform	YCbCr
	---- XMP ----
XMPToolkit	Adobe XMP Core 4.2-c020 1.124078, Tue Sep 11 2007 23:21:40
Urgency	1 (most urgent)
LegacyIPTCDigest	0AF9963ABE0645B9BBA74DCF1832AC7C
ColorMode	RGB
ICCProfileName	sRGB IEC61966-2.1
AuthorsPosition	Staff Photographer
DateCreated	2007:04:19 11:03:15-06:00
City	Watseka
State	Illinois
Country	United States
Headline	Farmer planting onions
CaptionWriter	Susan Brown
TransmissionReference	CSA farms
Instructions	Newspapers Out, Original Artixscan 4000 of color negative file, 160 ISO (frame 35a) is 7.6 x 11.2 at 500ppi, in Adobe RGB.
Credit	Big Newspaper
Source	John Doe / Big Newspaper
CreateDate	2005:02:01 00:30:44-06:00
ModifyDate	2010:07:14 03:38:26-05:00
MetadataDate	2010:07:16 03:24:33+01:00
CreatorTool	Adobe Photoshop CS5 Windows
Rating	3
Label	Red

DocumentID	adobe:docid:photoshop:2e336694-741a-11d9-aca3-a8f1d89cf0a9
InstanceID	xmp.iid:7914DC2B238FDF119A76B605C7D1F39B
OriginalDocumentID	adobe:docid:photoshop:2e336694-741a-11d9-aca3-a8f1d89cf0a9
Format	image/jpeg
IntellectualGenre	Profile
Location	Moore family farm
CountryCode	US
Marked	True
WebStatement	http://www.bignewspaper.com/
CountryCode	US
Prefs	Tagged:0, ColorClass:1, Rating:3, FrameNum:-00001
HasCrop	False
AlreadyApplied	True
MaxAvailWidth	5600
MaxAvailHeight	3800
PLUSVersion	1.2.0
ModelReleaseStatus	Limited or Incomplete Model Releases
MinorModelAgeDisclosure	Age 25 or Over
PropertyReleaseStatus	Limited or Incomplete Property Releases
ImageSupplierImageID	bng01661gda
ExifVersion	0221
ExifImageWidth	516
ExifImageHeight	350
ColorSpace	sRGB

Orientation	Horizontal (normal)
ImageWidth	516
ImageHeight	350
PhotometricInterpretation	RGB
SamplesPerPixel	3
XResolution	72
YResolution	72
ResolutionUnit	inches
DerivedFromInstanceID	xmp.iid:7814DC2B238FDF119A76B605C7D1F39B
DerivedFromDocumentID	adobe:docid:photoshop:2e336694-741a-11d9-aca3-a8f1d89cf0a9
DerivedFromOriginalDocumentID	adobe:docid:photoshop:2e336694-741a-11d9-aca3-a8f1d89cf0a9
HistoryAction	saved, saved, saved, saved, converted, derived, saved, saved, converted, derived, saved
HistoryInstanceID	xmp.iid:99A4E4000D21DF1192FCC7C8367E9136, xmp.iid:A52161322A41DF11BCE3BEC6B31A1DB2, xmp.iid:35FCB81F4B8EDF11B777A809894B1B88, xmp.iid:0D78DC8F218FDF119A76B605C7D1F39B, xmp.iid:0E78DC8F218FDF119A76B605C7D1F39B, xmp.iid:7814DC2B238FDF119A76B605C7D1F39B, xmp.iid:7914DC2B238FDF119A76B605C7D1F39B
HistoryWhen	2010:02:24 00:23:50-06:00, 2010:04:05 22:12:13-05:00, 2010:07:13 01:51:55-05:00, 2010:07:14 03:26:55-05:00, 2010:07:14 03:26:55-05:00, 2010:07:14 03:38:26-05:00, 2010:07:14 03:38:26-05:00
HistorySoftwareAgent	Adobe Photoshop CS4 Windows, Adobe Photoshop CS5 (12.0x20100323 [20100323.m.1172 2010/03/23:02:00:00 cutoff; m branch]) Windows, Adobe Photoshop CS5 Windows, Adobe Photoshop CS5 Windows, Adobe Photoshop CS5 Windows, Adobe Photoshop CS5 Windows, Adobe Photoshop CS5 Windows
HistoryChanged	/, /, /, /, /, /, /

HistoryParameters	from image/tiff to application/vnd.adobe.photoshop, converted from image/tiff to application/vnd.adobe.photoshop, from application/vnd.adobe.photoshop to image/jpeg, converted from application/vnd.adobe.photoshop to image/jpeg
Description	After digging the furrows another ten yards with the tractor, Jim Moore hops off to hand-set more leeks and onions.
Creator	John Doe
Subject	agriculture, farm laborer, farmer, field hand, field worker, humans, occupation, people, agricultural, agronomy, crops, onions, vegetable crops, plants, vegetables, outdoors, outside, agricultural equipment, tractor, gender, male, men
Title	01661gdx
Rights	©2010 Big Newspaper, all rights reserved
CreatorAddress	Big Newspaper, 123 Main Street
CreatorCity	Boston
CreatorRegion	Massachusetts
CreatorPostalCode	02134
CreatorCountry	United States
CreatorWorkTelephone	+1 (800) 1234567
CreatorWorkEmail	johndoe@bignewspaper.com
CreatorWorkURL	http://www.bignewspaper.com
Scene	011900
SubjectCode	04001000, 04001001
UsageTerms	For consideration only, no reproduction without prior permission
PersonInImage	Jim Moore
LocationCreatedSublocation	Moore family farm
LocationCreatedCity	Watseka

LocationCreatedProvinceState	Illinois
LocationCreatedCountryName	United States of America
LocationCreatedCountryCode	US
LocationCreatedWorldRegion	North America
LocationShownSublocation	Moore family farm
LocationShownCity	Watseka
LocationShownProvinceState	Illinois
LocationShownCountryName	United States
LocationShownCountryCode	US
LocationShownWorldRegion	North America
OrganisationInImageName	Prairieland Community Sponsored Agriculture
OrganisationInImageCode	http://www.prairielandcsa.org/
Event	Spring Planting
RegistryOrganisationID	http://www.plus-id.org
RegistryItemID	B01-9C8-7EC-65F
LicensorName	Big Newspaper Group
LicensorID	http://plus-id.org/99-G5-H2W
LicensorTelephone1	+1 (800) 1234567
LicensorEmail	info@bignewspaper.com
LicensorURL	http://www.bignewspapergroup.com/
ModelReleaseID	Bng20070419jd
PropertyReleaseID	Bng20070420jd
CopyrightOwnerName	Big Newspaper Group
CopyrightOwnerID	http://www.bignewspapergroup.com/
ImageCreatorName	John Doe
ImageCreatorID	99-G5-H2W

ImageSupplierName	Big Newspaper Group
ImageSupplierID	http://plus-id.org/77-B5-H2W
BitsPerSample	8, 8, 8
---- IPTC ----	
ApplicationRecordVersion	2
ObjectName	01661gdx
Urgency	1 (most urgent)
Keywords	agriculture, farm laborer, farmer, field hand, field worker, humans, occupation, people, agricultural, agronomy, crops, onions, vegetable crops, plants, vegetables, outdoors, outside, agricultural equipment, tractor, gender, male, men
SpecialInstructions	Newspapers Out, Original Artixscan 4000 of color negative file, 160 ISO (frame 35a) is 7.6 x 11.2 at 500ppi, in Adobe RGB.
DateCreated	2007:04:19
TimeCreated	11:03:15-06:00
By-line	John Doe
By-lineTitle	Staff Photographer
City	Watseka
Sub-location	Moore family farm
Province-State	Illinois
Country-PrimaryLocationName	United States
OriginalTransmissionReference	CSA farms
Headline	Farmer planting onions
Credit	Big Newspaper
Source	John Doe / Big Newspaper
CopyrightNotice	©2010 Big Newspaper, all rights reserved
Caption-Abstract	After digging the furrows another ten yards with the

	tractor, Jim Moore hops off to hand-set more leeks and onions.
Writer-Editor	Susan Brown
	---- Photoshop ----
IPTCDigest	0af9963abe0645b9bba74dcf1832ac7c
XResolution	72
DisplayedUnitsX	inches
YResolution	72
DisplayedUnitsY	inches
GlobalAngle	126
GlobalAltitude	30
CopyrightFlag	True
URL	http://www.bignewspaper.com/
PhotoshopThumbnail	(Binary data 5623 bytes, use -b option to extract)
PhotoshopQuality	10
PhotoshopFormat	Standard
ProgressiveScans	3 Scans

17 Appendix 6 – List of picture libraries surveyed

Adams Picture Library – <http://www.adamspicturelibrary.com/>

Africa Photos – <http://www.africaphotos.com/>

Agripicture – <http://www.agripicture.com/>

Alamy – <http://www.alamy.com/>

ArenaPAL – <http://www.arenapal.com/>

Art Resource – <http://www.artres.com/>

Artindustri – <http://www.artindustri.com/>

Big Stock Photo – <http://www.bigstockphoto.com/>

Bridgeman Art Library – <http://www.bridgeman.co.uk/>

cmpimages – <http://www.cmpimages.com/>

Chiro.org Image Archive – <http://www.chiro.org/chimages/>

Crash Picture Agency – <http://www.crashpa.net/>

Christie's Images – <http://www.christiesimages.com/>

Dennis Kunkel Science Stock Photography – <http://www.denniskunkel.com/>

DK Images – <http://www.dkimages.com/>

Dr Stock – <http://www.doctorstock.com/>

ePicScotland – <http://www.epicscotland.com/>

Eyevine – <http://www1.eyevinearchive.com/>

FLPA – <http://www.flpa-images.co.uk/>

Fotolia – <http://www.fotolia.co.uk/>

Fotosearch – <http://www.fotosearch.co.uk/>

FreeFoto.com – <http://www.freefoto.com/>

Garden and Wildlife Matters Photo Library – <http://www.gardenmatters.uk.com/>

Heritage Image Partnership – <http://www.heritage-images.com/>

ILN Picture Library – <http://www.ilnpictures.co.uk/>

Image After – <http://www.imageafter.com/>

Inmagine – <http://www.inmagine.com/>

iStockphoto – <http://www.istockphoto.com/>

Jupiter Images – <http://www.jupiterimages.com/>

Lightbox – <http://www.the-lightbox.com/>

Lonely Planet Images – <http://www.lonelyplanetimages.com>

Medical Photographic Library (MPL) – <http://medphoto.wellcome.ac.uk/>

Mediscan – <http://www.mediscan.co.uk/>

MirrorPix – <http://www.mirropix.com/>

Morgue File – <http://www.morguefile.com/>

Natural History Pictures – <http://www.natural-history-pictures.co.uk/>

Nucleus – <http://catalog.nucleusinc.com/>

Open Photo – <http://openphoto.net/>

Photo-key – <http://www.photo-key.com/>

Photo Researchers Stock Photography – <http://www.photoresearchers.com/>

Photos.com – <http://www.photos.com/en/>

Photos To Go – <http://www.photostogo.com/>

Robert Harding Picture Library – <http://www.robertharding.com/>

Science and Society Picture Library – <http://www.scienceandsociety.co.uk/>

Science Photo Library – <http://www.sciencephoto.com/index.html>

SeaPics – <http://www.seapics.com/>

Stock.xchng – <http://www.sxc.hu/>

Stockxpert – <http://www.stockxpert.com/>

UK Landscape Stock Photo Library – <http://www.buyimage.co.uk/>

UK Stock Images – <http://www.ukstockimages.com/>

Untitled Picture Library – <http://www.untitled.co.uk/>

View Buildings – <http://www.viewbuildings.com/>

Visuals Unlimited – <http://www.visualsunlimited.com/>

18 Appendix 7 – Sample Description of Implied Licence

From the *ONIX for Books: Product Information Format Implementation and Best Practice – Release 3.0*. (Bell, G., 2011).

Implied license to use ONIX data

When a data supplier provides an ONIX message to a data recipient, there is a clear invitation extended to the recipient with an 'implied license' to use the product data supplied for the purposes of cataloguing, trading in, merchandising, promoting and selling the products described, both internally within the recipient organization and in customer-facing applications. However, ONIX recipients should treat the implied license as non-transferable, and should not redistribute the data to other organizations either commercially or non-commercially without express permission. ONIX recipients should also treat compliance with Embargo, valid From/Until and Announcement dates, Audience limitations and the like as a duty of the implied licence, and ensure reasonable efforts are made to process data updates from the supplier in a timely way. A typical retailer might not require anything beyond this implied license.

Formal agreements such as would be required for commercial or non-commercial redistribution of data supplied in an ONIX message (e.g. by a data aggregator or through provision of third-party API access to the data) should be concluded separately between data supplier and recipient.

Certain data included within, cited by or linked from an ONIX message – for example, excerpts from the product, tables of contents, cover or contributor images and other resources – is subject to copyright. For the avoidance of doubt, it is strongly recommended that recipients wishing to make *modifications* to any data (as opposed to using the data as supplied) should secure an explicit license from the data supplier rather than relying solely on the implied license.

19 Appendix 8 – Sample DDEX messages

#	DDEX
1	<?xml version="1.0" encoding="UTF-16" ?>
2	<ern:NewReleaseMessage xmlns:ern="http://ddex.net/xml/2010/ern-main/312" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://ddex.net/xml/2010/ern-main/312 http://ddex.net/xml/2010/ern-main/312/ern-main.xsd" MessageSchemaVersionId="2010/ern-main/312">
3	<MessageHeader>
4	<MessageThreadId />
5	<MessageId />
6	<MessageSender>
7	<PartyId />
8	</MessageSender>
9	<MessageRecipient>
10	<PartyId />
11	</MessageRecipient>
12	<MessageCreatedDateTime>2001-01-01T10:00:00Z</MessageCreatedDateTime>
13	</MessageHeader>
14	<UpdateIndicator>OriginalMessage</UpdateIndicator>
15	<ResourceList>
16	<SoundRecording>
17	<SoundRecordingId>
18	<ISRC>JPBV00300425</ISRC>
19	</SoundRecordingId>
20	<ResourceReference>A1</ResourceReference>
21	<ReferenceTitle>
22	<TitleText LanguageAndScriptCode="ja-Jpan">ユートピアは見えてるのに</TitleText>
23	</ReferenceTitle>
24	<Duration>PT4H38M25S</Duration>
25	<SoundRecordingDetailsByTerritory>
26	<TerritoryCode>Worldwide</TerritoryCode>
27	<Title>
28	<TitleText LanguageAndScriptCode="ja-kana">ユートピアハミエテルノニ</TitleText>
29	</Title>
30	<Title>
31	<TitleText LanguageAndScriptCode="en-Latn">Utopia Ha Mieterunoni</TitleText>

32	</Title>
33	<DisplayArtist>
34	<PartyName>
35	<FullName LanguageAndScriptCode="en-Latn">DEEN</FullName>
36	</PartyName>
37	<ArtistRole>FeaturedArtist</ArtistRole>
38	</DisplayArtist>
39	<DisplayArtist>
40	<PartyName>
41	<FullName LanguageAndScriptCode="en-Latn">Deen</FullName>
42	</PartyName>
43	<ArtistRole>FeaturedArtist</ArtistRole>
44	</DisplayArtist>
45	<DisplayArtist>
46	<PartyName>
47	<FullName LanguageAndScriptCode="ja-kana">ディーン</FullName>
48	</PartyName>
49	<ArtistRole>FeaturedArtist</ArtistRole>
50	</DisplayArtist>
51	</SoundRecordingDetailsByTerritory>
52	</SoundRecording>
53	<SoundRecording>
54	<SoundRecordingId>
55	<ISRC>JPBV00300428</ISRC>
56	</SoundRecordingId>
57	<ResourceReference>A2</ResourceReference>
58	<ReferenceTitle>
59	<TitleText LanguageAndScriptCode="en-Latn">Re-Birth</TitleText>
60	<SubTitle LanguageAndScriptCode="ja-kana">インストゥルメンタル </SubTitle>
61	</ReferenceTitle>
62	<Duration>PT4H25M60S</Duration>
63	<SoundRecordingDetailsByTerritory>
64	<TerritoryCode>Worldwide</TerritoryCode>
65	<Title>
66	<TitleText LanguageAndScriptCode="ja-kana">リバース</TitleText>
67	<SubTitle LanguageAndScriptCode="ja-kana">インストゥルメンタル </SubTitle>
68	</Title>
69	<Title>
70	<TitleText LanguageAndScriptCode="en-Latn">Re-Birth</TitleText>
71	<SubTitle LanguageAndScriptCode="en-Latn">Instrumental</SubTitle>

72	</Title>
73	<DisplayArtist>
74	<PartyName>
75	<FullName LanguageAndScriptCode="en-Latn">DEEN</FullName>
76	</PartyName>
77	<ArtistRole>FeaturedArtist</ArtistRole>
78	</DisplayArtist>
79	<DisplayArtist>
80	<PartyName>
81	<FullName LanguageAndScriptCode="en-Latn">Deen</FullName>
82	</PartyName>
83	<ArtistRole>FeaturedArtist</ArtistRole>
84	</DisplayArtist>
85	<DisplayArtist>
86	<PartyName>
87	<FullName LanguageAndScriptCode="ja-kana">ディーン</FullName>
88	</PartyName>
89	<ArtistRole>FeaturedArtist</ArtistRole>
90	</DisplayArtist>
91	</SoundRecordingDetailsByTerritory>
92	</SoundRecording>
93	</ResourceList>
94	<ReleaseList>
95	<Release>
96	<ReleaseId>
97	<ICPN IsEan="false">82876556242</ICPN>
98	</ReleaseId>
99	<ReleaseId>
100	<ICPN IsEan="true">828765562423</ICPN>
101	</ReleaseId>
102	<ReferenceTitle>
103	<TitleText LanguageAndScriptCode="ja-Jpan">ユートピアは見えてるのに</TitleText>
104	</ReferenceTitle>
105	<ReleaseResourceReferenceList>
106	<ReleaseResourceReference>A1</ReleaseResourceReference>
107	<ReleaseResourceReference>A2</ReleaseResourceReference>
108	</ReleaseResourceReferenceList>
109	<ReleaseDetailsByTerritory>
110	<TerritoryCode>Worldwide</TerritoryCode>
111	<Title>

112	<TitleText LanguageAndScriptCode="ja-kana">ユートピアハミエテルノニ </TitleText>
113	</Title>
114	<Title>
115	<TitleText LanguageAndScriptCode="en-Latn">Utopia Ha Mieterunoni</TitleText>
116	</Title>
117	<DisplayArtist>
118	<PartyName>
119	<FullName LanguageAndScriptCode="ja-hani">DEEN</FullName>
120	</PartyName>
121	<ArtistRole>FeaturedArtist</ArtistRole>
122	</DisplayArtist>
123	<DisplayArtist>
124	<PartyName>
125	<FullName LanguageAndScriptCode="en-Latn">Deen</FullName>
126	</PartyName>
127	<ArtistRole>FeaturedArtist</ArtistRole>
128	</DisplayArtist>
129	<DisplayArtist>
130	<PartyName>
131	<FullName LanguageAndScriptCode="ja-kana">ディーン</FullName>
132	</PartyName>
133	<ArtistRole>FeaturedArtist</ArtistRole>
134	</DisplayArtist>
135	</ReleaseDetailsByTerritory>
136	</Release>
137	</ReleaseList>
138	</ern:NewReleaseMessage>

20 Appendix 9 – MovieLabs Common Metadata Schema (used in EIDR)

Sample xml for each data field in an EIDR record can be found in the EIDR Data Fields specification:
http://eidr.org/documents/EIDR-1.02_Data_Fields.pdf.

#	Common Schema
1	<?xml version="1.0" encoding="UTF-8" ?>
2	<xs:schema xmlns:md="http://www.movielabs.com/schema/md/v1.1/md" xmlns:xs="http://www.w3.org/2001/XMLSchema" targetNamespace="http://www.movielabs.com/schema/md/v1.1/md" elementFormDefault="qualified" attributeFormDefault="unqualified" version="1.1">
3	<xs:simpleType name="id-type">
4	<xs:restriction base="xs:anyURI" />
5	</xs:simpleType>
6	<xs:simpleType name="orgID-type">
7	<xs:restriction base="xs:string" />
8	</xs:simpleType>
9	<xs:simpleType name="ContentID-type">
10	<xs:restriction base="md:id-type" />
11	</xs:simpleType>
12	<xs:complexType name="ContentIdentifier-type">
13	<xs:sequence>
14	<xs:element name="Namespace" type="xs:string" />
15	<xs:element name="Identifier" type="xs:string" />
16	<xs:element name="Location" type="xs:anyURI" minOccurs="0" />
17	</xs:sequence>
18	</xs:complexType>
19	<xs:simpleType name="AssetPhysicalID-type">
20	<xs:restriction base="md:id-type" />
21	</xs:simpleType>
22	<xs:simpleType name="AssetLogicalID-type">
23	<xs:restriction base="md:id-type" />
24	</xs:simpleType>
25	<xs:complexType name="ContentRatingDetail-type">
26	<xs:sequence>
27	<xs:element name="Region" type="md:Region-type" />
28	<xs:element name="System" type="xs:string" />
29	<xs:element name="Value" type="xs:string" />
30	<xs:element name="Reason" type="xs:string" minOccurs="0" />

	maxOccurs="unbounded" />
31	<xs:element name="LinkToLogo" type="xs:anyURI" minOccurs="0" />
32	</xs:sequence>
33	</xs:complexType>
34	<xs:complexType name="ContentRating-type">
35	<xs:sequence>
36	<xs:choice>
37	<xs:element name="NotRated" type="xs:boolean" fixed="true" />
38	<xs:element name="Rating" type="md:ContentRatingDetail-type" maxOccurs="unbounded" />
39	</xs:choice>
40	<xs:element name="AdultContent" type="xs:boolean" minOccurs="0" />
41	</xs:sequence>
42	</xs:complexType>
43	<xs:complexType name="ContentSequenceInfo-type">
44	<xs:sequence>
45	<xs:element name="Number" type="xs:int" />
46	<xs:element name="HouseSequence" type="xs:string" minOccurs="0" />
47	</xs:sequence>
48	</xs:complexType>
49	<xs:complexType name="BasicMetadataParent-type">
50	<xs:choice>
51	<xs:element name="Parent" type="md:BasicMetadata-type" />
52	<xs:element name="ParentContentID" type="md:ContentID-type" />
53	</xs:choice>
54	<xs:attribute name="relationshipType">
55	<xs:simpleType>
56	<xs:restriction base="xs:string">
57	<xs:enumeration value="isclipof" />
58	<xs:enumeration value="isepisodeof" />
59	<xs:enumeration value="isseasonof" />
60	<xs:enumeration value="ispieceof" />
61	<xs:enumeration value="ispartof" />
62	<xs:enumeration value="isderivedfrom" />
63	<xs:enumeration value="iscompositeof" />
64	<xs:enumeration value="issupplementto" />
65	<xs:enumeration value="ispromotionfor" />
66	</xs:restriction>
67	</xs:simpleType>
68	</xs:attribute>
69	</xs:complexType>

70	<xs:complexType name="BasicMetadataInfo-type">
71	<xs:sequence>
72	<xs:element name="TitleDisplay19" type="xs:string" />
73	<xs:element name="TitleDisplay60" type="xs:string" minOccurs="0" />
74	<xs:element name="TitleSort" type="xs:string" />
75	<xs:element name="ArtReference" minOccurs="0" maxOccurs="unbounded">
76	<xs:complexType>
77	<xs:simpleContent>
78	<xs:extension base="xs:anyURI">
79	<xs:attribute name="resolution" type="xs:string" />
80	</xs:extension>
81	</xs:simpleContent>
82	</xs:complexType>
83	</xs:element>
84	<xs:element name="Summary190">
85	<xs:complexType>
86	<xs:simpleContent>
87	<xs:extension base="xs:string">
88	<xs:attribute name="cast" type="xs:boolean" />
89	</xs:extension>
90	</xs:simpleContent>
91	</xs:complexType>
92	</xs:element>
93	<xs:element name="Summary400" minOccurs="0">
94	<xs:complexType>
95	<xs:simpleContent>
96	<xs:extension base="xs:string">
97	<xs:attribute name="cast" type="xs:boolean" />
98	</xs:extension>
99	</xs:simpleContent>
100	</xs:complexType>
101	</xs:element>
102	<xs:element name="Summary4000" minOccurs="0">
103	<xs:complexType>
104	<xs:simpleContent>
105	<xs:extension base="xs:string">
106	<xs:attribute name="cast" type="xs:boolean" />
107	</xs:extension>
108	</xs:simpleContent>
109	</xs:complexType>
110	</xs:element>

111	<xs:element name="DisplayIndicators" minOccurs="0" maxOccurs="unbounded">
112	<xs:simpleType>
113	<xs:restriction base="xs:string">
114	<xs:enumeration value="CC" />
115	<xs:enumeration value="F" />
116	<xs:enumeration value="P" />
117	<xs:enumeration value="DD" />
118	<xs:enumeration value="SAP" />
119	<xs:enumeration value="DVS" />
120	</xs:restriction>
121	</xs:simpleType>
122	</xs:element>
123	<xs:element name="Genre" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
124	<xs:element name="Keyword" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
125	<xs:element name="VersionNotes" type="xs:string" minOccurs="0" />
126	<xs:element name="Region" type="md:Region-type" minOccurs="0" maxOccurs="unbounded" />
127	<xs:element name="OriginalTitle" type="xs:string" />
128	<xs:element name="CopyrightLine" type="xs:string" minOccurs="0" />
129	<xs:element name="PeopleLocal" type="md:BasicMetadataPeople-type" minOccurs="0" maxOccurs="unbounded" />
130	</xs:sequence>
131	<xs:attribute name="language" type="xs:language" use="required" />
132	<xs:attribute name="default" type="xs:boolean" />
133	</xs:complexType>
134	<xs:complexType name="BasicMetadataJob-type">
135	<xs:sequence>
136	<xs:element name="JobFunction">
137	<xs:complexType>
138	<xs:simpleContent>
139	<xs:extension base="md:Role-type">
140	<xs:attribute name="scheme" type="xs:string" />
141	</xs:extension>
142	</xs:simpleContent>
143	</xs:complexType>
144	</xs:element>
145	<xs:element name="JobDisplay" type="xs:string" minOccurs="0" />
146	<xs:element name="BillingBlockOrder" type="xs:int" minOccurs="0" />
147	<xs:element name="Character" type="xs:string" minOccurs="0" maxOccurs="unbounded" />

148	<xs:element name="Guest" type="xs:boolean" minOccurs="0" />
149	</xs:sequence>
150	</xs:complexType>
151	<xs:complexType name="BasicMetadataPeople-type">
152	<xs:sequence>
153	<xs:element name="Job" type="md:BasicMetadataJob-type" maxOccurs="unbounded" />
154	<xs:element name="Name" type="md:PersonName-type" />
155	<xs:element name="Identifier" type="md:PersonIdentifier-type" minOccurs="0" maxOccurs="unbounded" />
156	<xs:element name="Gender" minOccurs="0">
157	<xs:simpleType>
158	<xs:restriction base="xs:string">
159	<xs:enumeration value="male" />
160	<xs:enumeration value="female" />
161	<xs:enumeration value="neutral" />
162	<xs:enumeration value="plural" />
163	</xs:restriction>
164	</xs:simpleType>
165	</xs:element>
166	</xs:sequence>
167	</xs:complexType>
168	<xs:complexType name="BasicMetadata-type">
169	<xs:sequence>
170	<xs:element name="UpdateNum" minOccurs="0">
171	<xs:simpleType>
172	<xs:restriction base="xs:int">
173	<xs:minInclusive value="1" />
174	</xs:restriction>
175	</xs:simpleType>
176	</xs:element>
177	<xs:element name="LocalizedInfo" type="md:BasicMetadataInfo-type" maxOccurs="unbounded" />
178	<xs:element name="RunLength" type="xs:duration" minOccurs="0" />
179	<xs:element name="ReleaseYear" type="xs:gYear" />
180	<xs:choice>
181	<xs:element name="ReleaseDate" type="xs:date" minOccurs="0" />
182	<xs:element name="ReleaseDateTime" type="xs:dateTime" minOccurs="0" />
183	</xs:choice>
184	<xs:element name="ReleaseHistory" type="md:ReleaseHistory-type" minOccurs="0" maxOccurs="unbounded" />
185	<xs:element name="WorkType" type="xs:string" />

186	<xs:element name="PictureColorType" type="md:ColorType-type" minOccurs="0" />
187	<xs:element name="PictureFormat" type="xs:string" minOccurs="0" />
188	<xs:element name="ThreeD" type="xs:boolean" minOccurs="0" />
189	<xs:element name="AltIdentifier" type="md:ContentIdentifier-type" minOccurs="0" maxOccurs="unbounded" />
190	<xs:element name="RatingSet" type="md:ContentRating-type" minOccurs="0" />
191	<xs:element name="People" type="md:BasicMetadataPeople-type" minOccurs="0" maxOccurs="unbounded" />
192	<xs:element name="CountryOfOrigin" type="md:Region-type" minOccurs="0" />
193	<xs:element name="AssociatedOrg" minOccurs="0" maxOccurs="unbounded">
194	<xs:complexType>
195	<xs:complexContent>
196	<xs:extension base="md:OrgName-type">
197	<xs:attribute name="role" type="xs:string" />
198	</xs:extension>
199	</xs:complexContent>
200	</xs:complexType>
201	</xs:element>
202	<xs:element name="SequenceInfo" type="md:ContentSequenceInfo-type" minOccurs="0" />
203	<xs:element name="Parent" type="md:BasicMetadataParent-type" minOccurs="0" maxOccurs="unbounded" />
204	</xs:sequence>
205	<xs:attribute name="ContentID" type="md:ContentID-type" use="required" />
206	</xs:complexType>
207	<xs:complexType name="AbbreviatedMetadataInfo-type">
208	<xs:sequence>
209	<xs:element name="TitleBrief" type="xs:string" />
210	<xs:element name="ArtReference" type="xs:anyURI" minOccurs="0" maxOccurs="unbounded" />
211	<xs:element name="SummaryShort" type="xs:string" />
212	<xs:element name="DisplayIndicators" minOccurs="0" maxOccurs="unbounded">
213	<xs:simpleType>
214	<xs:restriction base="xs:string">
215	<xs:enumeration value="CC" />
216	<xs:enumeration value="F" />
217	<xs:enumeration value="P" />
218	<xs:enumeration value="DD" />
219	<xs:enumeration value="SAP" />
220	<xs:enumeration value="DVS" />
221	</xs:restriction>

222	</xs:simpleType>
223	</xs:element>
224	</xs:sequence>
225	<xs:attribute name="language" type="xs:language" use="required" />
226	<xs:attribute name="default" type="xs:boolean" />
227	</xs:complexType>
228	<xs:complexType name="AbbreviatedMetadata-type">
229	<xs:sequence>
230	<xs:element name="UpdateNum" minOccurs="0">
231	<xs:simpleType>
232	<xs:restriction base="xs:int">
233	<xs:minInclusive value="1" />
234	</xs:restriction>
235	</xs:simpleType>
236	</xs:element>
237	<xs:element name="LocalizedInfo" type="md:AbbreviatedMetadataInfo-type" maxOccurs="unbounded" />
238	<xs:element name="Rating" type="md:ContentRating-type" minOccurs="0" maxOccurs="unbounded" />
239	<xs:element name="AltIdentifier" type="md:ContentIdentifier-type" minOccurs="0" maxOccurs="unbounded" />
240	<xs:element name="Studio" type="xs:string">
241	<xs:annotation>
242	<xs:documentation>Equivalent to DisplayName</xs:documentation>
243	</xs:annotation>
244	</xs:element>
245	</xs:sequence>
246	<xs:attribute name="ContentID" type="md:ContentID-type" use="required" />
247	</xs:complexType>
248	<xs:complexType name="DigitalAssetAudioEncoding-type">
249	<xs:sequence>
250	<xs:element name="Codec" type="xs:string" />
251	<xs:element name="CodecType" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
252	<xs:element name="BitrateMax" type="xs:integer" minOccurs="0" />
253	<xs:element name="SampleRate" type="xs:integer" minOccurs="0" />
254	<xs:element name="SampleBitDepth" type="xs:integer" minOccurs="0" />
255	</xs:sequence>
256	</xs:complexType>
257	<xs:complexType name="DigitalAssetAudioData-type">
258	<xs:sequence>
259	<xs:element name="Description" type="xs:string" minOccurs="0" />

260	<xs:element name="Type" type="xs:string" minOccurs="0" />
261	<xs:element name="Encoding" type="md:DigitalAssetAudioEncoding-type" minOccurs="0" />
262	<xs:element name="Language">
263	<xs:complexType>
264	<xs:simpleContent>
265	<xs:extension base="xs:language">
266	<xs:attribute name="dubbed" type="xs:boolean" />
267	</xs:extension>
268	</xs:simpleContent>
269	</xs:complexType>
270	</xs:element>
271	<xs:element name="Channels" type="xs:string" />
272	<xs:element name="TrackReference" type="xs:string" minOccurs="0" />
273	</xs:sequence>
274	</xs:complexType>
275	<xs:complexType name="DigitalAssetVideoEncoding-type">
276	<xs:sequence>
277	<xs:element name="Codec" type="xs:string" />
278	<xs:element name="CodecType" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
279	<xs:element name="MPEGProfile" type="xs:string" minOccurs="0" />
280	<xs:element name="MPEGLLevel" type="xs:string" minOccurs="0" />
281	<xs:element name="BitrateMax" type="xs:integer" minOccurs="0" />
282	</xs:sequence>
283	</xs:complexType>
284	<xs:complexType name="DigitalAssetVideoPicture-type">
285	<xs:sequence>
286	<xs:element name="AspectRatio" type="xs:string" />
287	<xs:element name="PixelAspect" minOccurs="0">
288	<xs:simpleType>
289	<xs:restriction base="xs:string">
290	<xs:enumeration value="NTSC" />
291	<xs:enumeration value="PAL" />
292	<xs:enumeration value="square" />
293	<xs:enumeration value="other" />
294	</xs:restriction>
295	</xs:simpleType>
296	</xs:element>
297	<xs:element name="WidthPixels" type="xs:int" minOccurs="0" />
298	<xs:element name="HeightPixels" type="xs:int" minOccurs="0" />
299	<xs:element name="ActiveWidthPixels" type="xs:int" minOccurs="0" />

300	<xs:element name="ActiveHeightPixels" type="xs:int" minOccurs="0" />
301	<xs:element name="FrameRate" type="xs:int" minOccurs="0" />
302	<xs:element name="Progressive" type="xs:boolean" minOccurs="0" />
303	<xs:element name="Type3D" type="xs:boolean" minOccurs="0" />
304	</xs:sequence>
305	</xs:complexType>
306	<xs:complexType name="DigitalAssetImageData-type">
307	<xs:sequence>
308	<xs:element name="Width" type="xs:int" />
309	<xs:element name="Height" type="xs:int" />
310	<xs:element name="Encoding" type="xs:string" />
311	<xs:element name="TrackReference" type="xs:string" minOccurs="0" />
312	</xs:sequence>
313	</xs:complexType>
314	<xs:complexType name="DigitalAssetVideoData-type">
315	<xs:sequence>
316	<xs:element name="Description" type="xs:string" minOccurs="0" />
317	<xs:element name="Type" type="xs:string" minOccurs="0" />
318	<xs:element name="Encoding" type="md:DigitalAssetVideoEncoding-type" minOccurs="0" />
319	<xs:element name="Picture" type="md:DigitalAssetVideoPicture-type" />
320	<xs:element name="ColorType" type="md:ColorType-type">
321	<xs:annotation>
322	<xs:documentation>BW, Color, Colorized, etc.</xs:documentation>
323	</xs:annotation>
324	</xs:element>
325	<xs:element name="SubtitleLanguage" minOccurs="0">
326	<xs:complexType>
327	<xs:simpleContent>
328	<xs:extension base="xs:language">
329	<xs:attribute name="closed" type="xs:boolean" />
330	</xs:extension>
331	</xs:simpleContent>
332	</xs:complexType>
333	</xs:element>
334	<xs:element name="TrackReference" type="xs:string" minOccurs="0" />
335	</xs:sequence>
336	</xs:complexType>
337	<xs:complexType name="DigitalAssetSubtitleData-type">
338	<xs:sequence>
339	<xs:element name="Format">

340	<xs:complexType>
341	<xs:simpleContent>
342	<xs:extension base="xs:string">
343	<xs:attribute name="SDImage" type="xs:boolean" />
344	<xs:attribute name="HDImage" type="xs:boolean" />
345	</xs:extension>
346	</xs:simpleContent>
347	</xs:complexType>
348	</xs:element>
349	<xs:element name="Description" type="xs:string" minOccurs="0" />
350	<xs:element name="Type" type="xs:string" minOccurs="0" />
351	<xs:element name="FormatType" type="xs:string" minOccurs="0" />
352	<xs:element name="Language" type="xs:language" />
353	<xs:element name="TrackReference" type="xs:string" minOccurs="0" />
354	</xs:sequence>
355	</xs:complexType>
356	<xs:complexType name="DigitalAssetInteractiveData-type">
357	<xs:sequence>
358	<xs:element name="Type" type="xs:string" />
359	<xs:element name="Language" type="xs:language" />
360	<xs:element name="TrackReference" type="xs:string" minOccurs="0" />
361	</xs:sequence>
362	</xs:complexType>
363	<xs:complexType name="DigitalAssetMetadata-type">
364	<xs:choice>
365	<xs:element name="Audio" type="md:DigitalAssetAudioData-type" />
366	<xs:element name="Video" type="md:DigitalAssetVideoData-type" />
367	<xs:element name="Subtitle" type="md:DigitalAssetSubtitleData-type" />
368	<xs:element name="Image" type="md:DigitalAssetImageData-type" />
369	<xs:element name="Interactive" type="md:DigitalAssetInteractiveData-type" />
370	</xs:choice>
371	</xs:complexType>
372	<xs:simpleType name="ContainerProfile-type">
373	<xs:restriction base="xs:string" />
374	</xs:simpleType>
375	<xs:complexType name="ContainerSpecific-type">
376	<xs:sequence>
377	<xs:any processContents="lax" minOccurs="0" maxOccurs="unbounded" />
378	</xs:sequence>
379	</xs:complexType>

380	<xs:simpleType name="ContainerType-type">
381	<xs:restriction base="xs:string" />
382	</xs:simpleType>
383	<xs:complexType name="ContainerMetadata-type">
384	<xs:sequence>
385	<xs:element name="Type" type="md:ContainerType-type" minOccurs="0" />
386	<xs:element name="Track" type="md:DigitalAssetMetadata-type" maxOccurs="unbounded" />
387	<xs:element name="ContainerSpecificMetadata" type="md:ContainerSpecific-type" minOccurs="0" />
388	</xs:sequence>
389	</xs:complexType>
390	<xs:simpleType name="CompObjID-type">
391	<xs:restriction base="md:id-type" />
392	</xs:simpleType>
393	<xs:complexType name="CompObjEntry-type">
394	<xs:sequence>
395	<xs:element name="DisplayName" minOccurs="0">
396	<xs:complexType>
397	<xs:simpleContent>
398	<xs:extension base="xs:string">
399	<xs:attribute name="language" type="xs:language" />
400	</xs:extension>
401	</xs:simpleContent>
402	</xs:complexType>
403	</xs:element>
404	<xs:element name="Entry" type="md:CompObjEntry-type" minOccurs="0" maxOccurs="unbounded" />
405	<xs:choice>
406	<xs:element name="ContentID" type="md:ContentID-type" />
407	<xs:element name="BasicMetadata" type="md:BasicMetadata-type" />
408	</xs:choice>
409	</xs:sequence>
410	</xs:complexType>
411	<xs:complexType name="CompObjData-type">
412	<xs:complexContent>
413	<xs:extension base="md:CompObj-type">
414	<xs:sequence>
415	<xs:element name="DisplayName" minOccurs="0" maxOccurs="unbounded">
416	<xs:complexType>
417	<xs:simpleContent>
418	<xs:extension base="xs:string">

419	<xs:attribute name="language" type="xs:language" />
420	</xs:extension>
421	</xs:simpleContent>
422	</xs:complexType>
423	</xs:element>
424	</xs:sequence>
425	<xs:attribute name="CompObjID" type="md:CompObjID-type" />
426	</xs:extension>
427	</xs:complexContent>
428	</xs:complexType>
429	<xs:complexType name="CompObj-type">
430	<xs:sequence>
431	<xs:element name="Entry" type="md:CompObjEntry-type" maxOccurs="unbounded" />
432	</xs:sequence>
433	</xs:complexType>
434	<xs:complexType name="ReleaseHistory-type">
435	<xs:sequence>
436	<xs:element name="ReleaseType" type="xs:string" />
437	<xs:element name="DistrTerritory" type="md:Region-type" minOccurs="0" />
438	<xs:element name="DateTime" minOccurs="0">
439	<xs:complexType>
440	<xs:simpleContent>
441	<xs:extension base="xs:dateTime">
442	<xs:attribute name="scheduled" type="xs:boolean" />
443	</xs:extension>
444	</xs:simpleContent>
445	</xs:complexType>
446	</xs:element>
447	<xs:element name="Date">
448	<xs:complexType>
449	<xs:simpleContent>
450	<xs:extension base="xs:date">
451	<xs:attribute name="scheduled" type="xs:boolean" />
452	</xs:extension>
453	</xs:simpleContent>
454	</xs:complexType>
455	</xs:element>
456	<xs:element name="Description" type="xs:string" minOccurs="0" />
457	<xs:element name="ReleaseOrg" type="md:OrgName-type" minOccurs="0" maxOccurs="unbounded" />
458	</xs:sequence>

459	</xs:complexType>
460	<xs:simpleType name="Role-type">
461	<xs:restriction base="xs:string" />
462	</xs:simpleType>
463	<xs:complexType name="NVPair-type">
464	<xs:sequence>
465	<xs:element name="Name" type="xs:string" />
466	<xs:element name="Value" type="xs:string" />
467	</xs:sequence>
468	</xs:complexType>
469	<xs:simpleType name="ColorType-type">
470	<xs:restriction base="xs:string">
471	<xs:enumeration value="color" />
472	<xs:enumeration value="bandw" />
473	<xs:enumeration value="colorized" />
474	<xs:enumeration value="composite" />
475	</xs:restriction>
476	</xs:simpleType>
477	<xs:complexType name="Region-type">
478	<xs:choice>
479	<xs:element name="country">
480	<xs:simpleType>
481	<xs:restriction base="xs:string">
482	<xs:pattern value="[A-Z][A-Z]" />
483	</xs:restriction>
484	</xs:simpleType>
485	</xs:element>
486	<xs:element name="countryRegion">
487	<xs:simpleType>
488	<xs:restriction base="xs:string">
489	<xs:pattern value="[A-Z][A-Z]-[A-Z0-9]+" />
490	</xs:restriction>
491	</xs:simpleType>
492	</xs:element>
493	</xs:choice>
494	</xs:complexType>
495	<xs:complexType name="OrgName-type">
496	<xs:sequence>
497	<xs:element name="DisplayName" type="xs:string" />
498	<xs:element name="SortName" type="xs:string" minOccurs="0" />
499	</xs:sequence>

500	<xs:attribute name="organizationID" type="md:orgID-type" />
501	</xs:complexType>
502	<xs:complexType name="PersonName-type">
503	<xs:sequence>
504	<xs:element name="DisplayName" type="xs:string" />
505	<xs:element name="SortName" type="xs:string" minOccurs="0" />
506	<xs:element name="FirstGivenName" type="xs:string" minOccurs="0" />
507	<xs:element name="SecondGivenName" type="xs:string" minOccurs="0" />
508	<xs:element name="FamilyName" type="xs:string" minOccurs="0" />
509	<xs:element name="Suffix" type="xs:string" minOccurs="0" />
510	<xs:element name="Moniker" type="xs:string" minOccurs="0" />
511	</xs:sequence>
512	</xs:complexType>
513	<xs:complexType name="ContactInfo-type">
514	<xs:sequence>
515	<xs:element name="Name" type="xs:string" />
516	<xs:element name="PrimaryEmail" type="xs:string" />
517	<xs:element name="AlternateEmail" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
518	<xs:element name="Address" type="xs:string" minOccurs="0" maxOccurs="unbounded" />
519	<xs:element name="Phone" minOccurs="0" maxOccurs="unbounded">
520	<xs:complexType>
521	<xs:simpleContent>
522	<xs:extension base="xs:string">
523	<xs:attribute name="type" type="xs:string" />
524	</xs:extension>
525	</xs:simpleContent>
526	</xs:complexType>
527	</xs:element>
528	</xs:sequence>
529	</xs:complexType>
530	<xs:complexType name="PersonIdentifier-type">
531	<xs:sequence>
532	<xs:element name="Identifier" type="xs:string" />
533	<xs:element name="Namespace" type="xs:string" />
534	<xs:element name="ReferenceLocation" type="xs:anyURI" minOccurs="0" />
535	</xs:sequence>
536	</xs:complexType>
537	<xs:complexType name="DateTimeRange-type">
538	<xs:sequence>
539	<xs:element name="Start" type="xs:dateTime" />

540	<xs:element name="End" type="xs:dateTime" />
541	</xs:sequence>
542	</xs:complexType>
543	</xs:schema>