



The LCC Framework for the Rights Data Supply Chain

v1.0, April 2013

The Linked Content Coalition (LCC) is a cross-media, multinational coalition of more than 40 partners from the media and creative industries, including representatives of authors and artists, working together with their standards bodies to establish automated communications between rightsholders and those who wish to use content. FAQs about LCC can be found [here](#).

1 Introduction

The principal goal of the Linked Content Coalition is to enable **greater legitimate use of digital content** through better management of data relating to rights across the network. The LCC recognises that securing the highest possible level of automation in licensing will reduce barriers to entry, reduce cost in the supply chain, increase volume of use and encourage innovation.

The LCC's first outputs in this task are the documents which comprise the **LCC Framework for the Rights Data Supply Chain**, introduced in this paper. These documents introduce and describe:

- the **[LCC Rights Reference Model \(RRM\)](#)**, a formal, general and extensible reference data model for representing intellectual property rights and entitlements for any media or content, and
- the "best practice" principles for **[Identification](#)**, **[Messaging](#)** and how users access rights data and acquire licenses.

The **next steps** are the first implementations of the LCC Framework, outlined below.

1.1 Problem and solution

There is a huge and rapidly expanding amount of digital content available through the internet, but it is often difficult for those (companies or individuals) who want to trade in rights to find each other. Transactions are too people-heavy and therefore both expensive and inefficient. New business opportunities are deferred because of the cost of complying

with multiple standards. Multiple registries and exchanges are unable to communicate effectively or automatically.

This context for the LCC Framework and the direction of the future work of the LCC will be expanded in a White Paper (to follow), which will describe the elements of the multimedia rights data landscape as the LCC understands it today.

The underlying principle of LCC, set out in the earliest LCC documents, is that "the answer to the machine is the machine."¹ If digital rights data are complex, and getting more so by the day, then the network itself holds the key to putting these in better order: digital technology, working with *well-structured, machine-interpretable data*.

The LCC Framework is an innovative design for integrating data in this network. It encompasses rights in any kind of content and media and describes how parties in the network may support a seamless, cross-media rights trading infrastructure in which rights owners may be discovered, requests and offers made and where licenses are available.

1.3 What the LCC is not

The LCC is neutral in relation to specific legal constraints. Laws, agreements and policies, and the values which they may attach to specific rights, are for lawmakers and rightsholders to determine. The LCC's role is to be able to represent the full range of their possible choices in simple-as-possible machine-readable ways as the data moves up and down the supply chain, and then to do all it can to encourage solutions to manage this data. The Framework must therefore accommodate rights granted under different jurisdictions, and entitlements (such as the role of administering licences or collecting royalties) which are assigned by agreement or policy and not by law.

The LCC is also neutral about business models. Rights may be exploited for value or exercised freely, according to the wishes of whoever has the right to assign them. Distinctions between "commercial" and "cultural" domains are also not relevant at this generic level: libraries, museums and archives may be acquirers and exploiters of rights and creators of content just as commercial publishers or aggregators are, as can a bank or a supermarket or a private individual loading a video or photograph onto a file-sharing service. Other apparent "anomalies" such as works in the public domain, fair use provisions or "orphan works" are also to be viewed not as exceptional cases but as other kinds of rights which should be expressible consistently alongside others.

¹ The original aims and methodology of stage 1 of the LCC are set out in the [LCC Project Plan](#).

Nor does the LCC replace any existing standards or registries: just the opposite. Its task is to design and implement ways in which existing data, identifier and messaging standards can work together (or *interoperate*) in the most seamless and automated ways. Only where suitable tools or standards which are needed do not exist does the LCC encourage new initiatives to fill gaps.

1.4 Radical use of best practise

All of this requires a somewhat radical approach to the understanding of "rights" in terms of scope and data architecture. That is what the LCC Framework takes, but it does so while making use of existing best practise in its methods. The result is reassuringly coherent. Rights administration will always involve complexity because the content, the uses to which it is put and the conditions which apply really are complex, and have become much more so in the digital age. The Framework, however, offers an integrated strategy which can be applied both "top down" and "bottom up", making use of existing schemas and infrastructure but describing ways of creating, aggregating and transforming complex, multimedia data to fill gaps in the network.

The LCC vision is not Utopian: there will always be problems with the creation and supply of data, and there will always be many rights transactions and data flows which cannot be wholly automated through data weaknesses or the need for human negotiation. However, there is a great deal that can be done to create a data network suited to digital needs and capabilities that puts users and providers in contact as seamlessly as possible. It has been achieved for much online commerce, and there is no reason why it cannot now be done for rights.

2 The LCC Framework documents

The first stage of LCC set out to define:

- the types of things that occupy the network, and their relationships (**2.1 Rights Reference Model**);
- how to identify things in the network (**2.2 Principles of Identification**);
- how the rights data passes through the network (**2.3 Principles of Messaging**);
- how users access the data (**2.4 Principles of User Interface**).

2.1 The LCC Rights Reference Model (RRM)

The **LCC Rights Reference Model (RRM)** is a formal, general and extensible reference data model for representing intellectual property rights and entitlements for any media or content. The RRM identifies eight types of entity, their core attributes and the relationships between them. One very significant feature is the single **Right** entity, through which permissions and entitlements of all kinds (including ownership claims and permissions) may be expressed to any level of detail with their associated prohibitions and conditions. This provides a breadth and a simplicity to the model which other rights-based models typically lack.

[*The LCC Rights Reference Model v1.0.pdf*](#)

The RRM has been created using a "building block" model, the **LCC Entity Model**, which gives a general structure for the attributes of any kind of Entity. The underlying Entity Model is what allows the RRM to be "specialised" to any level of detail and configured to meet the requirements of different domains within a single model by the addition of specialised vocabulary and rules. It also allows for other types of Entity to be added to the model if its scope is extended in future.

[*The LCC Entity Model v1.0.pdf*](#)

For convenience the RRM is also expressed as two UML (Unified Modelling Language) class models in an EAP (Enterprise Architect) file, and the UML class models are also published as JPG images.

[*The LCC Rights Reference Model v1.0.eap*](#)

[*The LCC Rights Reference Model v1.0.jpg*](#)

[*The LCC Entity Model v1.0.jpg*](#)

2.1.1 RRM Use Cases

The RRM is one of a family of models and standards based on, or consistent with, the <indec> Metadata Framework² (2000), which includes the DDEX and ONIX message standards from the music and text publishing sectors. The RRM must, by definition, be semantically richer than any schema to be mapped into it, so is expected to cover at the

² *The <indec> metadata framework Version 2.0, June 2000*

very least all the rights messages in the rights schemas identified in the table in the Appendix to this paper.

The RRM has been validated by mapping a range of use cases into it, chosen to cover a wide range of types of rights data, including from schemas in the Appendix table. This process of validation will be continued in a formal way within the RDI project³.

[LCC RRM Exemplary Use Cases - 18 separate pdfs](#)

2.1.2 The Common Rights Format (CRF)

The RRM is an abstract data model which can be expressed or implemented in many different syntactic forms (such as a relational database schema, an object model, an RDF or XML schema or a formal ontology) without changing its meaning. As an example, the **Common Rights Format (CRF)** is provided as an XML schema (XSD) which enables data to be expressed in XML in a way that complies with the RRM. The CRF, or some variant of it, will be used in the RDI project referred to in the next steps below. Like many XSDs it is a set of several (in this case, three) dependent schemas:

[LCCCommonRightsFormat.xsd](#)

[LCCEntityModel.xsd](#)

[LCCAllowedValues.xsd](#)

A sample record is provided, showing Use Case R1 in an XML document compliant with the CRF:

[CRF Sample RRM Use Case R1.xml](#)

The CRF is not proposed as a general Rights message in the form published here, which is more suitable to support a "hub" aggregation and transformation process of the kind described in the **Interoperability** section below. Specific schemas and messages for more specialised use, or to "plug gaps" in the network, could be created as profiles of the CRF.

³ It had been intended to map examples of each of the listed schemas in the first stage of LCC, but limited resources meant that some of this work has been deferred until the RDI project. Those mapped (ODRL and PLUS) were chosen for their breadth and/or complexity. From informal review and knowledge of the other listed schemas it is not anticipated that there will be any significant issues raised in mapping.

2.2 LCC Principles of Identification

With the RRM having defined the key eight kinds of entity which populate the network, the **LCC Principles of Identification** document contains the LCC recommendations for the way in which these entities are best identified. Eight recommendations are presented as a model of best practise for supporting the highest level of automation, trust and accuracy within the supply chain and network, followed by a review of the status quo for different entity types and sectors.

[*LCC Principles of Identification v1.0.pdf*](#)

2.3 LCC Principles of Messaging

Data and identifiers pass through a data flow called in RRM the **Rights Data Supply Chain**. The **LCC Principles of Messaging** defines the three essential roles in this chain, and provides an analysis of the Information flows which move along it. The document describes the basic messaging architecture and generic Message requirements which may be used to specify message formats, message exchange protocols and choreographies.

[*LCC Principles of Messaging v1.0.pdf*](#)

2.4 LCC Principles of User Interface

How, finally, does a potential content user access the data in the network? The final workstream of the first stage of LCC has been deferred to the next stage, in which principles for user interface and possibly some proposed standards for service types and user iconography will be defined.

[*LCC Principles of User Interface - not yet published.*](#)

3 Next steps: implementations of the LCC Framework

With much of the Framework completed the next stage of the LCC will concentrate on implementation. There are two main areas in which LCC expects to concentrate next to accelerate the development of the Digital Identifier Network for rights data:

- **Interoperability:** aggregation and transformation services using the RRM

- **Data capture:** Identification of parties, content and rights "at the point of entry" to the Web

There is also some remaining work to be done on the Framework:

- **LCC Principles of User Interface**
- Matrix of existing schemas and identifiers: LCC hopes to expand Table 1 above into a more comprehensive directory of metadata, identifier and message standards for the Digital Identifier Network.

As the initial LCC project is complete, the next phase of work will require a new form of governance for the LCC as a **standards consortium**.

3.1 LCC future: a standards consortium

It is proposed that the LCC will become a **consortium of standards bodies and registries**. Its members will be organisations who create and manage identifier, metadata and messaging standards for content of all media types. Any other organisation or person may show their support for the aims of the LCC by declaring themselves as an affiliate.

The core aim of the LCC will be to **facilitate and expand the legitimate use of content** in the digital network through the use of interoperable identifier and metadata standards.

The LCC will support **interoperability** between the computer systems of any and all legitimate participants in the digital network, including creators, rightsholders, publishers, aggregators, rights and content exchanges, retailers, consumers, cultural institutions, (including libraries, museums and archives) and their agents and associations. It will focus principally on rights-related data. The LCC will support the legitimate use of copyright, public domain and "orphan" works, under any business model or none, including "free use".

The LCC will not compete with the activities of existing standards bodies, but will deal with matters of common interest across existing standards bodies. It will also provide a forum to address matters which do not fall within the remit of others.

3.2 Interoperability: data aggregation and transformation services using the RRM

The first focus for LCC in 2013 is on **implementations of the RRM** for interoperability. In this approach a "hub" is provided which receives, transforms, aggregates (where necessary) and outputs queries and content and rights metadata between parties in the Rights Data Supply Chain. Other parties in the process do not need to deal with

The first of these is the **RDI** (Rights Data Integration project), an EC-funded initiative due to start in May 2013 in which established Sources and Exchanges from all major content sectors will create a prototype multi-media rights data network. An RRM-based "hub" will carry out the transformation of data to enable queries and responses to pass through the supply chain from potential Users to Sources and back, via Exchanges if needed.

The aim of RDI is to prove the value of an RRM-based hub for removing data barriers which inhibit the supply chain. Participants also expect that new commercial relationships will be established through RDI.

The UK Copyright Hub initiative also supports the LCC, and the RRM is being considered as the start point for its data architecture. Other potential "hub" implementations of RRM are in discussion.

The RDI project will result in the mapping of many **controlled vocabularies** to one another, from standardised and proprietary schemes. The project will make use of the Vocabulary Mapping Framework, which enables many-to-many mapping. The intent is that this work will provide the basis of a permanent set of LCC-authorized and maintained vocabulary mappings available to the network.

The consequences of aggregation-based interoperability are discussed more fully in the LCC white paper The Digital Identifier Network.

3.3 Data capture: identification "at the point of entry" to the Web

The second focus for LCC is expected to be on developing standards for **declaring rights and content data at the point of entry** to the Web.

As described in the Digital Identifier Network paper, a huge volume of works of all types (text, image, audio, audiovisual) is now being created and published directly on the Web, a significant proportion of it with potential commercial value and cultural value. In LCC this is described as "**self-published**" content, created and made available by people or organisations who do not operate as commercial publishers or distributors (although a good deal of conventional commercial content also makes use of the same services now). It includes what is often known as "user-generated content" (UGC).

Most of the services into which this content is loaded (including blogs, social media and content-sharing services such as YouTube, Flickr, Vimeo or Soundcloud) provide no means for the creators and publishers, content or rights metadata to be identified or declared in

the simple standard ways required to support the Digital Identifier Network: billions of new "orphans" are in effect being created daily as they enter the Network.

LCC's goal is that anyone loading a new work into the Web should have "at the point of entry to the Web" the option to be able to identify themselves (**Parties**), their content (**Creations**) and its **Rights** using standard or "shared" identifiers and vocabularies, and for that data to be declared in appropriate Sources, thus capturing authoritative data at source. There is no technological barrier to such services, and the RRM provides a data architecture to enable the widest range of diversity which may be required for different content and media: a service might use an existing service "off the shelf" or customise it to meet its own needs.

This approach and its potential benefits are discussed in more detail in the LCC white paper The Digital Identifier Network.

3.4 The vocabulary challenge

Many coming into contact with LCC have assumed that its metadata work would first and foremost be concerned with the *vocabulary* of rights, particularly the verbs (such as *play, copy, view, assign, stream, download, digitise* etc) which describe permitted or prohibited acts, and which vary in scope and meaning from one domain to another.

However, while these and other vocabularies are of great importance, they have been a secondary concern so far. The priority has been to develop an extensible data framework *within which* the authoritative mapping of such vocabularies can be managed (and maintained) so that existing schemes can operate together and new schemes can fit with them.

Such mapping work, and the development of a detailed ontology which will extend the RRM to enable it to embrace these diverse vocabularies, will take place in the RDI and other initiatives. A good deal of the necessary work has already been done on this in existing standards work including the Vocabulary Mapping Framework. It is likely that the LCC will make some direct contributions to standard metadata vocabularies in due course, but this will only come as a "gap-filling" exercise and as agreed by the community of standards which LCC represents.

Appendix

Table 1: Existing Rights metadata schemas⁴

Schema	Governance	Types of data (main content type, if any)
ccREL	Creative Commons	Rights Assignments
Common Works registration (CWR) message	CISAC	Control Rights (musical works)
Creative Barcode, Rights Reserved service	Creative Barcode	Control Rights, Policies
DDEX Work Licensing and Release Delivery messages	DDEX	Control Rights, Usage Rights (musical works, sound recordings)
METSRights	Digital Library Federation	Control Rights and Usage Rights (bibliographic)
ODRL 2.0 (various profiles)	W3C	Rights Assignments, Usage Rights
ONIX, various messages	EDItEUR	Control Rights and Usage Rights (books, journals)
OPDS Catalog Format v1.1	Open Publication Distribution System	Usage Rights (epublications)
PLUS Coalition License Data Format	PLUS Coalition	Rights Agreements, Usage Rights (still images)
RightsML 1.0 (profile of ODRL)	IPTC	Rights Agreements, Usage Rights (news)
XrML (various profiles including MPEG21 REL)	Contentguard/ MPEG-21	Rights Agreements, Usage Rights

⁴ This list is not exhaustive.