



FINAL REPORT OF THE RIGHTS DATA INTEGRATION PROJECT

Interim Version: May 2016

This version is complete except for some details of recommendations and exploitation plans in Section 4 below, which are still the subject of discussions between RDI partners and the Linked Content Coalition. When agreement on these are reached (hopefully in July 2016) the final version of this report will be published.

RDI Partners



ALBUM



PRO DUCENT
FORENINGEN



EPC

European
Publishers
Council



gettyimages®



m EDRA



RightsDirect rightscom

TABLE OF CONTENTS

1. Overview.....	4
2. Background.....	5
3. The copyright supply chain.....	6
4. The RDI Project.....	8
3.1. Creating standardised data	9
3.1.1. Range and expressiveness of data	10
3.1.2. Enriching data	10
3.2. Integrating data at the DI Hub.....	12
3.2.1. Data integration	12
3.2.2. Data queries	12
3.2.3. Scalability	13
3.3. Exchanges.....	14
3.3.1. Aggregator.....	14
3.3.2. General queries	15
3.3.3. Content recognition	16
3.3.4. Direct access.....	17
3.4. Dissemination activities.....	17
3.5. Results	19
4. Recommendations and future plans	19
4.1. Recommendations.....	20
4.2. Future plans.....	20
5. Contact details.....	21
5.1. Project contact details.....	21
5.2. Partner contact details	21

1. OVERVIEW

The problem is simple enough to describe: when you find a piece of content on the internet, it is often difficult and many times impossible to find out who own rights in it, whether it is free to be used, or how you can get a license to use it. It is a problem for creators, owners and users, and creates significant barriers to the legitimate use of content for all. The huge explosion in the volume of content now being published on the Web means this becomes more of an issue every day.

The RDI project has successfully demonstrated, with prototype systems, that there is a technical solution to this problem, based on the framework model created by the Linked Content Coalition (LCC), which can lead to a highly-automated, standards-based network for rights data without requiring participants to throw out their existing ways of working.

This solution serves the interests of creators, rights holders and users alike, as it is neutral in respect of business models: public domain and other freely usable content are supported just as well as copyright works, as is the individual creator/publisher or the multinational corporation or collective management organisation.

The project showed that rights information for any content or media type, and in any data format, can be automatically transformed and integrated in a common form so that it can be queried or passed on using standard protocols. It developed prototype **mapping, transformation, repository** and **query** tools based on the LCC model, and took complex rights data of all kinds – both standard and proprietary - from more than 20 different sources to show how this intermediary format could enable the declaration and discovery of rights and permissions data in a way that will be highly automatable and extensible.

Specific streams of the project showed how the approach integrates successfully with identification technology, in this case with content recognition software and the Digital Object Identifier (DOI).

The RDI project has produced a set of tools and prototypes which the project partners and others, including the LCC and the Copyright Hub, now expect to take forward into production systems. With widespread implementation it promises a dramatic improvement in the infrastructure of the digital rights network, removing barriers to trade and enabling many more creators and users to participate efficiently and legally.

2. BACKGROUND

The first decade and a half of the 21st century has seen an unprecedented change in the content and media domains – typically characterised as the text, image, music and audiovisual sectors. In that short time we have moved from a world in which the creation and publication of content was generally the preserve of a professional elite to a world in which more or less everyone with access to a computer (which is now about half the world's population) has become a creator and re-user of content.

The volume of content published daily on the internet is huge: one of the authors of this report has estimated, based on published statistics of web content, that the number of new and adapted works made public *every day* on the Web may now be greater than the total analogue output of civilisation since the invention of the printing press. It is a reasonable assumption that volume and diversity will only continue to increase.

But there is a serious, and growing, problem. The massive change in the availability of content has not been matched by a corresponding change in the information infrastructure that supports it. Physical supply chains have been replaced by digital, in which *data* is the key to distribution, access and administration of rights, but the network of content is developing much faster than the network of data that is needed to manage it effectively.

The Rights Data Integration (RDI) project (2013-15) was co-funded by the media sector and the European Union, to create a working prototype version of the framework for managing rights data previously specified by the Linked Content Coalition (LCC)¹, a not-for-profit, all-media consortium of global data standards bodies and others across the media/content sectors. LCC itself was a response to Commissioner Neelie Kroes' call in 2010 for "Big Ideas for the Digital Agenda".

The objective of LCC is to help establish more efficient communication of copyright information to those who wish to use content, to enable much greater legitimate use of content across the digital network. RDI was an essential step in the process to demonstrate that LCC's proposed solution to this problem is viable.

Through copyright, society values the role of every creator – the author and composer, the photographer and musician, the individual artist and the multinational corporation – and grants them the right to decide how their creations are used. The internet *promises* an end to traditional ways of doing business, an end to high barriers to entry and incumbency rights – but it also threatens an end to copyright as we know it – to the detriment of those who rely on it for their living and to society as a whole.

¹ <http://www.linkedcontentcoalition.org> The LCC Project was an unincorporated, 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. From this project came LCC Ltd, a company incorporated in England and established by the major media standards bodies to take on the work of the LCC Project.

Copyright has traditionally been managed through people-heavy processes – individual lawyer-crafted licences, communications on paper and managed in individual media silos, and more recently human beings searching and trying to interpret information on the internet. This is unsustainable in a digital age where multi-media is the norm and where everyone can create and publish their works to the world.

We need ways of managing copyright on the internet which are low cost, efficient and reduce barriers to entry: better ways to manage the digital rights supply chain in the interests of everyone in it. LCC has specified how that supply chain might be constructed: RDI proved that such solutions are practical.

3. THE COPYRIGHT SUPPLY CHAIN

Copyright is a commodity, and digital commodity supply chains critically rely on two things: unique **identifiers** and **schemas**.

Identifiers enable **automation** – computers require them for accurate processing. Every significant thing in a supply chain typically needs an identifier - not only the commodities themselves (whether those are books, washing machines, airline tickets, stock trades or bananas), but things related to the process: websites, customers, features and measurements of the commodity, orders, payments, credit cards, passwords, locations, telephone numbers, delivery events and so on. These identifiers may or may not be public, but they must be there.

But being able to identify a banana, where it came from, where it's going to, packaging dimensions, weight, and any other element in the banana supply chain is useless unless your computer also knows *how those identifiers relate to each other*: what may be called a **schema**. Published schemas, often expressed in APIs, are essential tools for interoperability. They enable people to develop software which can query and process all kinds of complex data efficiently.

The more that common identifiers and schemas are used in supply chains, the more automated activity there can be, the more barriers to entry can be broken down and trade increased to the benefit of providers and users alike. In stark contrast, mistakes or failures will arise whenever there is a break or error in the network (everything from forgetting your password at the beginning to having the wrong delivery address at the end).

In the copyright supply chain the core commodity that is traded, the equivalent of the banana or washing machine, is a right. For example, we don't trade the creation *The Lord of the Rings*, we trade the *right* to do something with a copy of that creation. The fatal flaw in today's copyright supply chain is that there is no way to identify the fundamental building block of that supply chain - the **right** itself.

LCC identified four key types of thing in a rights supply chain:

1. A **creation** or work in which rights exist or are being granted.
2. A **rightsholder** or rights user ("parties" in LCC)

3. A **rights assignment** - the basis on which a right to do something is established: at its most basic this may be the operation of the Berne Convention but includes any license, agreement, convention or policy which grants or withholds permission
4. A **right**² itself - that is, the entitlement that someone has or may acquire to do something with the creation

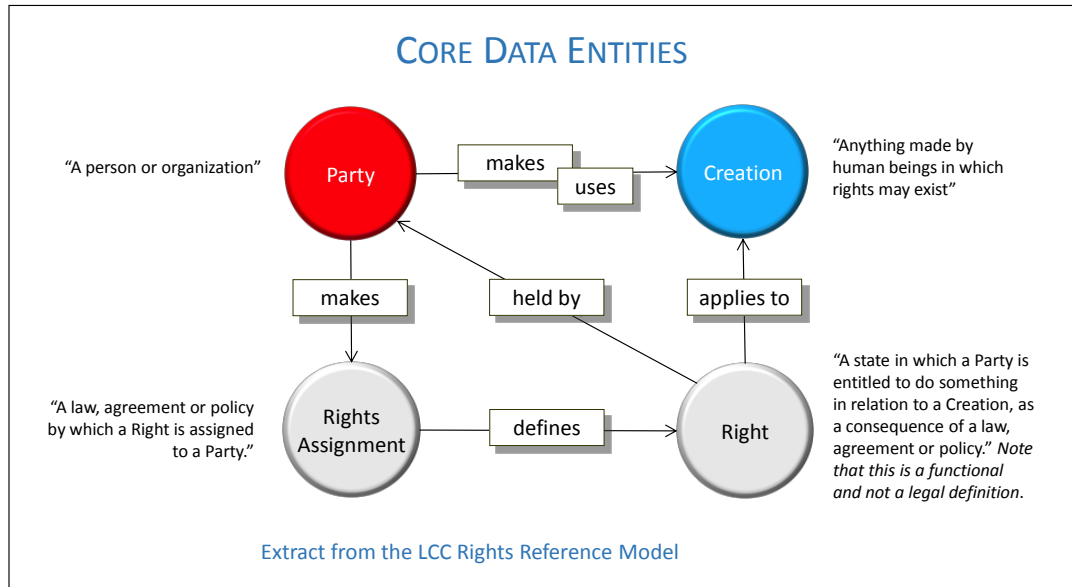


Diagram: the four main LCC model Entities, each of which requires an Identifier for effective rights supply chain processing.

Any missing identifier, or link between identifiers, causes disruption which will be all too familiar to participants in the supply chain, but aren't always analysed in this way. For example, orphan works are caused by not knowing the rightsholder Party for a Creation. Payments to rightsholders are delayed if the link between Creation and Party is not automatable. Creations aren't used if it isn't easy to find who holds the rights and what rights might be on offer. Cross border licensing is more complex if you can't identify the Party seeking to use rights and know what those rights are in a particular place and at a particular time.

To make matters worse, many of the additional data within the Right such as the permitted actions (e.g. "copy") or purposes (e.g. "educational use") or territories (e.g. "Europe excluding Germany") may not mean the same thing in one schema as they do in another. Without some means of establishing their relationship, they cannot be automatically processed together, although some may be very effective in particular sector supply chains.

² In LCC is "right" is defined very broadly as shown in the diagram. This definition includes copyright and other intellectual property rights, but in principle may cover any other kind of relevant permission, such as the right to collect money for the use or sale of content. This is a functional, not a legal, definition. The LCC considered using a different word (such as "entitlement") for this, but any similar word has its own misleading connotations, so we simply note that the label itself (in this case "right") is not important, but the definition which applies is. LCC and RDI have nothing to say about the validity of rights themselves: they are only interested in representing the claims that others may make about the entitlements they have.

We can summarise the rights data problem in three points:

- much basic rights data is not accessible at all;
- where it is accessible, it is in a variety of often complicated and incompatible schemas; and
- the basic commodity in the supply chain - *the right itself* - is not identified at all.

Identifier and data standards hold the key, but it is not practical or desirable to replace existing metadata standards designed for individual media sectors with a single, all-encompassing common standard. Instead the LCC Framework specified the essential framework for cross-media rights management and interoperability based on establishing a common **ontology** - a *data rulebook* including the generic LCC Rights Reference Model - which allows different parties to establish the relationships between different terms and schemas in a consistent and predictable fashion.

RDI took the LCC Framework specification and proved that it was technically viable to implement it.

4. THE RDI PROJECT

The aim of RDI was to prototype an implementation of the LCC Framework published in 2013³

*"to demonstrate how participants in the content supply chain can manage and trade rights for any and all types of usage across any and all types of content (physical, digital or abstract) in any and all media under any (or no) commercial model, and to support the provision of information to Users, some of which lead to the securing of licenses, some of which may be automated."*⁴

The core idea is simple. Someone has rights to a piece of content. Someone else wishes to use that content in a particular way, and needs to discover who it is who can give permission for that (or to discover if the content is in the public domain, or already freely available under a Creative Commons or other licence or policy).

However, the reality is often not as simple as that. There may be several rightsholders for different rights or territories, and these may change over time. There may be several items of content within another item, with different rightsholders (for example, texts and photographs in a book, or songs in a sound recording). More fundamentally, users want this process to be automated, so that (for example) a potential user of an image or text can do the equivalent of a right-click operation to discover the licensor of the work and go to the website automatically; or a picture researcher (for example) looking for rights for dozens of photographs can submit a single data file for querying and get back a complete set of results in a single, coherent form.

³ <http://www.linkedcontentcoalition.org/index.php/rights-data-network/lcc-framework>

⁴ From the RDI project proposal.

The LCC approach is that whether expressing rights or asking questions about rights both parties should use whatever language works best for them: neither party should impose its preferred language on the other. But for this to work there needs to be some behind the scenes *translation* to ensure that both parties know what they *mean* as they talk to each other. The LCC Framework describes how this can be achieved, and RDI demonstrates some ways in which it can be implemented.

RDI was designed to represent a simplified but very broad rights supply chain. Data was drawn from 20 sources⁵ including content providers/publishers/owners, collective management organisations, data standards bodies, a national library and a national archive and covered text, music, audiovisual and image content. RDI took that data and:

- 1 created standardised data,
- 2 integrated the standardised data in a fashion which allowed it to be queried, and
- 3 demonstrated four ways in which the standardised data might be used.

3.1. CREATING STANDARDISED DATA



There are two steps in this process – **mapping** and **transformation**. These are commonplace in systems everywhere, but what makes them different in RDI is the richness of the underlying schema to which a schema is mapped – the **Common Rights Format** or CRF.

Mapping is a one-off process to establish the correspondence between the elements in a schema and the CRF. The mapping generates a set of rules used by the transformation tool (see below) that can be used for any data conforming to the schema. Only when there are changes to the schema does the mapping itself need to be updated and new rules generated. In RDI the mapping work was undertaken by Rightscom Ltd.

The RDI data model: CRF

The common model underlying RDI is the LCC Rights Reference Model, which itself is a profile of the more generic LCC Entity Model. These abstract models are expressed for computer processing in a formal scheme called the Common Rights Format or CRF. Data in RDI is converted in and out of the CRF, which was proved to be a rich enough model to support the detailed semantics of all schemas provided.

All mappings are supported by, and in turn may add to, the underlying **ontology** which defines the terms used in any schema and then ensures that the approach taken in each mapping is consistent. It is expected that the RDI Ontology will be the basis of a permanent **LCC Ontology** in the future.

The **transformation tool** developed by CINECA uses the mapping rules to convert data from different sources into the Common Rights Format, and vice versa. This tool was based on an

⁵ Listed in section 5 below

existing open source tool (R2R) which was significantly enhanced by CINECA to meet RDI's requirements.

The prototype **mapping and ontology management tool** met all RDI's requirements, but a more robust and user-friendly tool is needed to run an operational service, for which the RDI tool serves as a full specification. The **transformation tool** also dealt with all of RDI's requirements well, but it was not specifically designed to be scalable for high volume or speed in production, so more work is also required to turn this into an operational tool. Some deficiency in the handling of variables in the original R2R tool used as the base also still requires re-engineering for a production model.

Innovations

The mapping and transformation work was innovative in a number of ways, enabling complex conditional mapping and a highly templated method of semantic mapping. The transformation tool was taken well beyond its original capability, and at the end a "standalone" transformation tool was demonstrated which used the CRF as an intermediate format in real time transformation, without storing any data permanently for querying.

3.1.1 RANGE AND EXPRESSIVENESS OF DATA

A wide range of use cases were successfully mapped, transformed, stored and queried in the common format. A number of widely used standard schemas for rights data, including **ONIX messages, DDEX messages, ODRL** and **Creative Commons** licenses were transformed into CRF, as well as a range of different kinds of proprietary schema.

A lot of non-rights data was also transformed in RDI, including bibliographic or descriptive data for creations, contact information for parties, and information supporting royalty distribution, demonstrating that the LCC Framework supports transformation of data of many kinds, and not just that relating to rights. It matters that metadata is not arbitrarily divided up because searches and queries over content and rights are often carried out together, and attribute of an item of content may feature in a rights constraint. RDI proved that the LCC model would support the representation of practically any metadata of interest to the content sector.

Range of schemas in RDI

More than 20 different schemas were mapped, some quite small and others very large and complex, covering these kinds of rights data among others:

- rightsholder statements
- licence offers for usage
- licence offers including delegating the right to license
- license offers with complex conditions
- author agreements with publishers
- completed licence agreements
- chains of agreements and rights, including reversions
- rights in repertoires or catalogues of content
- inheritance of terms and conditions to multiple rights

3.1.2 ENRICHING DATA

Much rights data is sparse, even non-existent. RDI demonstrated that mapping and transformation can add very significant **semantic gain**, making hidden meaning *explicit* so that it can then be queried effectively. This should be a major benefit in a network where so much rights data is either absent or implicit. Source data was enriched in five main ways described below:

3.1.2.1 ASSERTED RIGHTS

Parties often know what rights apply to their content, but don't make it public in a machine-readable, queryable way. At its simplest this implicit data might be enriched to a basic rights statements such as "*Party A has the Right to Licence Creation B exclusively for all rights anywhere in the World from Time T until an unknown time in the future*". This might be called a "digital © notice". It grants no permission to use, but provides the necessary signposting to allow a potential user to automatically identify and be directed to the rightsholder and potentially to enter into negotiations to acquire the right to use⁶.

3.1.2.2 TEMPLATES

Where a party uses an existing licence data schema, the transformation process can add the formal details of that licence data to descriptive data to generate extensive rights data. In RDI Creative Commons (CC) licenses were mapped to CRF and then this template applied to data from one source which used a range of CC licenses. The same licence template could then be applied to data from any source using the same CC licence.

A natural extension of this approach could be applied to dealing with different copyright laws in different jurisdictions. A law, convention or statute relating to copyright is, in terms of the LCC model, just another rights assignment, and can in principle be mapped to CRF in the same way as a licence offer. If this were done, then legal frameworks of any kind could be queried alongside other rights assignments or rights, and possible conflicts between them, either at a general level or relating specifically to individual works, could be detectable by routine queries.

For clarity: neither LCC nor RDI has anything to say about the content of the laws or statutes themselves, just as it has nothing to say about the terms expressed in a commercial contract. The LCC Framework is neutral regarding any legal or business model and concerned only with the consistent expression of those terms.

3.1.2.3 EXPANDING COMPRESSED DATA

Schemas commonly contain elements which summarise two or more pieces of information in a single element. The mapping process separates out the separate meanings and the relationship into its most expressive form to support querying.

3.1.2.4 CHAINS OF RIGHTS

⁶ In the LCC Ten Targets this is referred to as a Digital Rightsholder Statement.

One of the most challenging aspects of rights data are chains of rights or rights assignments. Rights pass by agreement from authors to publishers, Collective Management Organisations, sub publishers and others, typically in what are referred to as "repertoires" or "catalogues" of many items, which are sets of contents or rights which themselves may be added to or taken away from regularly. In two use cases, complex chains of this kind were presented in abbreviated forms, and these were expanded out into the CRF.

3.1.2.5 DERIVING SPECIFIC RIGHTS DATA FROM CLASS DATA

Where rights data is held at a catalogue/repertoire level the mapping process can generate individual rights data for each component item.

3.2 INTEGRATING DATA AT THE RDI HUB



In an operational system, once data has been transformed into CRF it might be stored in any number of databases in the network. For testing purposes in RDI most of the data was held in a single repository developed by CINECA, known as the **RDI Hub**. In a commercial model CINECA would be acting as a service provider to the sources. Data in the RDI Hub was stored in RDF, the Semantic Web Linked Data triple format, but although CRF is well suited to being implemented in RDF, CRF-based repositories could also be built using SQL (see 3.3.1 below), blockchain or in principle any other database architecture.

As a prototype, the repository did not have a number of functions which would be essential in a production system, such as automatic schema and data validation and a full audit trail. All data loads and transformation processes were manually triggered and where multiple data records existed for the same item, no attempt was made to check and de-duplicate or report on conflicts in individual attributes.

3.2.1 DATA INTEGRATION

Eight use cases provided data which contained information on Creations with identifiers which were also provided by other sources and the RDF repository automatically joined these together so that rights data from multiple sources could be viewed together. Indeed, the semantic merging of information relies on identifiers, implemented using URIs.

3.2.2 DATA QUERIES

The RDI queries, developed and tested at the RDI Hub, are the point where the main potential pay-off of the whole LCC Framework starts to be seen. Having a relatively simple generic data model like CRF means that it is possible to use the same query patterns (the same computer code) to discover similar information across different media and rights types.

For example, the question "who is a potential licensor for this song for performing rights in Europe today" and "who was licensing re-publication rights for this photograph for non-commercial use in Canada in 2014" should be answerable using the same algorithm: only specific values like "Canada", "Europe", "non-commercial", date ranges and so on will be variables, and these should all be terms which can be picked from the RDI Ontology, or values entered (like dates and quantities) in standard forms.

This is a very significant development: instead of any number of specialized applications querying any number of different databases for the same kinds of information in different forms, it offers the possibility of services using relatively simple common APIs and query patterns to be able to harvest, aggregate and interpret a very wide range of complex rights data. When combined with the data integration described above, there is great potential for querying efficiently across a wide range of data in very diverse forms in different repositories.

Finally CINECA demonstrated how data can be retrieved in real time from a Source, in this case Ediser's ERRO system (see 3.3.1 below), using a *pull* rather than the *push* method common to most of the use cases, and then integrated with data already in the Hub repository.

The RDI test queries by no means exhausted the potential of the LCC Framework, but were sufficient to demonstrate that it was capable of integrating diverse, all-media data into a single form which can be extensively queried for a range of different discovery and other functions.

RDI Query patterns

This was successfully tested in RDI using four main generic queries

- find creation
- find licensor
- find licence terms
- find creations available under a specific licence.

For prototype purposes a limited range of filters were chosen (including territory, right type, medium, context, user type, period start, period end, and identifier type) but in principle these can be extended to cover any type of query.

3.2.3 SCALABILITY

Scalability was not in scope for RDI. The very expressive nature of the CRF poses particular challenges in terms of the number of database reads which may be needed in any particular query path. Helpfully for RDI, the Copyright Hub⁷ was able to do some work on optimisation and testing of scalability, working in consultation with RDI.

The Copyright Hub initiative was not formally part of the RDI project, but has been developing in parallel with the same goals, and from the outset has been committed to using the LCC Framework and not duplicating RDI activity. The Copyright Hub has focussed so far on creating open source software to support licensing systems which can be accessed

⁷ The Copyright Hub is a UK-based initiative which aims to make the licensing of content easier. Further information may be found at <http://www.copyrighthub.co.uk/>

through Exchange services like those described above (a further example of which the Copyright Hub has itself developed). Its activities are therefore complementary to RDI, and it is expected that it will make use of the RDI Ontology and mapping services once they pass to permanent LCC governance.

Working with Rightscom, the Copyright Hub tested an optimised version of the CRF which significantly reduced the number of reads required for queries like the RDI queries. The Copyright Hub benchmarking tests showed a very significant improvement in query times, reaching acceptable levels for high volumes. The goal is to improve performance by flattening some of the elements, while preserving its full expressivity. The flattening of the model implemented in databases will put some more complexity into the Ontology and query patterns, but at a manageable level.

3.3 EXCHANGES

Four Exchanges (Ediser, RightsDirect, CIF and mEDRA) demonstrated the query capabilities of the LCC model, and then ways in which the LCC Framework can be integrated with other common and developing technologies.



3.3.1. AGGREGATOR

Ediser played the most varied role in RDI, as a Source aggregating data from several other Sources, and creating a CRF-based Exchange using traditional technologies which interacted with other Exchanges and the RDI Hub. **ERRO** was developed to support Ediser in its rights management operations and unlike the other systems in RDI was designed to be an ongoing operational system.

ERRO might be described as an operational RDI in microcosm, and provides a prototype for a relatively lightweight in-house implementation of the LCC Framework. Rights and descriptive data come from its own Sources in a mix of standard formats (ONIX schemas) and proprietary formats, and are transformed into a CRF-based database which supports all kinds of processing. ERRO enables publishers to configure a wide variety of different and complex licensing offers for their publications, and these offers are then made available to other Exchanges (RightsDirect in the RDI project) and to the RDI Hub, and can be accessed in ERRO by the mEDRA rights-aware DOI.

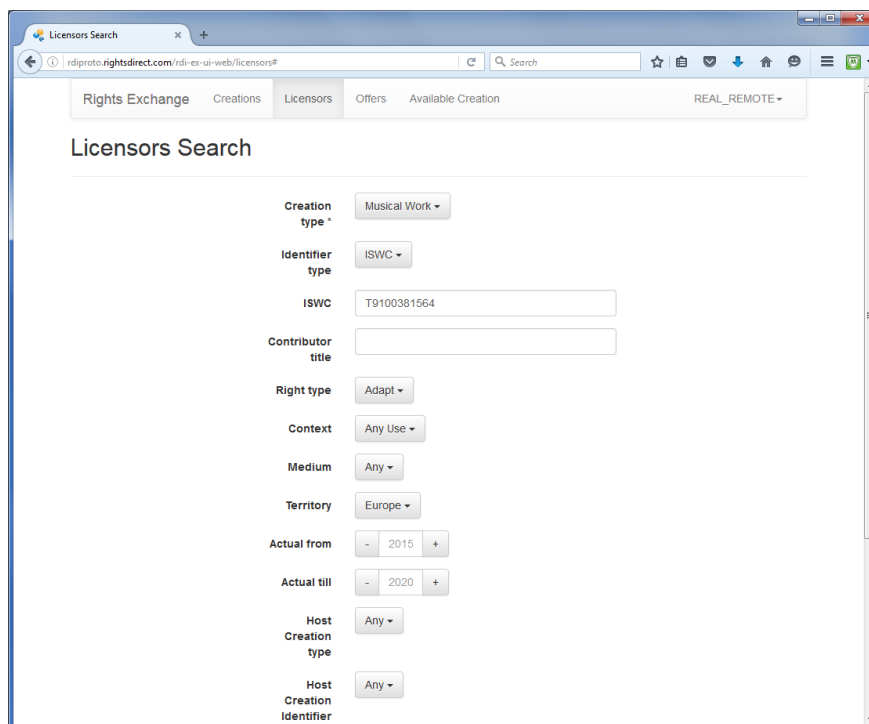
ERRO supports the use of CRF as a data interchange format so the RightsDirect Exchange is able to retrieve data from ERRO just as it can from the RDI Hub. ERRO also chose to use RDI Ontology terms for its controlled vocabularies so that it could export data using LCC/RDI terms and avoid the need for further translation into the RDI Hub.

ERRO is not a full CRF database implementation, as it used the CRF only as a reference model and implemented a flattened version optimised for use in a smaller scale, more traditional database development.

3.3.2. GENERAL QUERIES

The query capability at the Hub provided the platform for the prototype RDI general Exchange, developed by RightsDirect. RightsDirect already offers a range of commercial licensing services primarily for text-based publications (books and journals). Their RDI Exchange development was to test the potential of the LCC Framework for offering such services across all media types.

The RightsDirect Exchange executes the same queries as CINECA, but remotely, retrieving answers in real time from the Hub and processing and presenting them locally for action (which in a production environment could then lead to the acquisition of licenses). The Exchange service was also used to interact directly with another Exchange system (mEDRA) and with a Source (Ediser), by-passing the Hub but using the CRF for further real-time query and interchange. This is described under CRF-based Aggregator (Ediser ERRO system) above.



The screenshot shows a web browser window titled "Licensors Search" with the URL "rdiproto.rightsdirect.com/rdi-ex-ui-web/licensors?". The page has a navigation bar with "Rights Exchange", "Creations", "Licensors", "Offers", and "Available Creation". Below the navigation bar, the "Licensors Search" form is displayed with the following fields:

- Creation type: Musical Work
- Identifier type: ISWC
- ISWC: T9100381564
- Contributor title: (empty)
- Right type: Adapt
- Context: Any Use
- Medium: Any
- Territory: Europe
- Actual from: 2015
- Actual till: 2020
- Host Creation type: Any
- Host Creation Identifier: Any

RightsDirect developed a user-friendly Exchange interface which enabled users to make the full range of RDI queries, with drop-down lists from which the allowed values in different categories could be chosen.

There are three points of interaction between the remote RightsDirect Exchange and the Hub and Ontology:

First, the allowed values for the drop-down lists were provided in a data feed from the RDI Ontology, which was updated when new options were added to the Ontology. In RDI the feed and updates were manually triggered but in a production system they would be

automated, like other routine software updates. In an operational environment, ensuring that information of this kind remains synchronised across the network to be used by multiple Exchanges is essential.

Second, the variable values for each query were sent from the Exchange to the RDI Hub in an agreed format. They did not send the complete query, only the data values to be deployed in the query engine at the RDI Hub, so the communication format is relatively simple –no complicated query logic.

Third, query results were returned from the RDI Hub to the Exchange in a format of the Exchange's choice. Here there was an additional benefit for RDI, as RightsDirect chose to use CRF and thereby test it as an interchange schema.

These three interfaces established and tested a relatively simple, flexible and easily configurable protocol for remote query/discovery services.



3.3.3. CONTENT RECOGNITION

Content recognition technology is a very important method of identifying digital Creations. The CEPIC Image Finder (CIF) Exchange was developed to show how this technology could interact with rights data in general and with the LCC Framework in particular, and to show how multiple hubs could be integrated.

Queries in CIF are started simply by loading a photographic image into a browser page. The CIF connects to different photographic library Sources (Getty Images, AGE and Album in this project, any number in a production system), which already employ image recognition software (in RDI, Getty used its own Picscout technology, and Album and AGE used a system developed by Album).

The results of matches of the image searches are sent back to the CIF browser (there may be multiple matches, as a particular photograph is commonly licensed non-exclusively through several libraries), along with basic data about the potential licensor and the scope of rights available. This is integrated and displayed in the CIF Exchange interface, enabling users to click on a link to the licensor's website or service.

CIF has then established two-way communication with the RDI Hub. Rights data gathered from Sources through CIF can be sent to the RDI Hub, being transformed on the way from the CIF schema into CRF, like any other RDI Source data. In the other direction, CIF can query the RDI Hub using any image identifier provided by its Sources, and retrieve any rights data at the Hub, coming from non-CIF sources, for that image. The RDI Hub data is displayed along with the CIF data to the CIF user. CIF therefore can use a combination of image recognition and conventional identifiers for a wholly automated and real-time query.

ALBUM

Most of the technical work for this workstream was carried out by Album Archivo Fotografico.



3.3.4. DIRECT ACCESS

mEDRA demonstrated how rights data can be accessed from anywhere using an identifier, extending the existing DOI (Digital Object Identifier⁸) resolution capability to make a "rights-aware DOI". With this, users anywhere on the Web may be offered a range of options to discover rights (and other) data and services. The service can be activated just by clicking on an icon or link on a webpage, and presents the user with whatever service choices are on offer, and a further click then accesses the service.

The mEDRA service demonstrated how the CRF can enable rights data integration between two independent organisations *without* having to go through an RDI Hub. The service was used in conjunction with both the RightsDirect Exchange and the ERRO system using the combination of the DOI multiple resolution technology with the CRF to enable real-time discovery of rights data.

The rights-aware DOI is an implementation of what the DOI community calls "contextual resolution", and this specific application is being promoted within the International DOI Foundation for general use.

3.4 DISSEMINATION ACTIVITIES



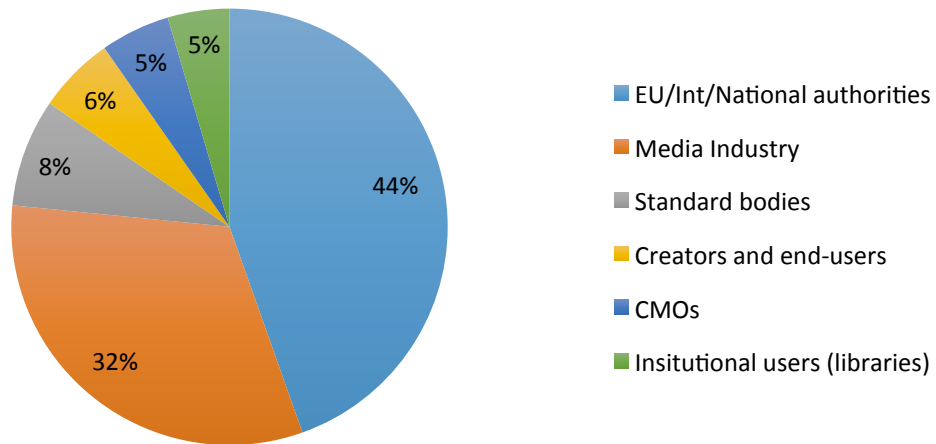
The main goal of the dissemination activities in RDI has been to maintain the momentum created by the LCC.

The dissemination work package, led by Europe Analytica, and supported by the European Writers Council created a web site and communication materials (fact sheets, presentations, newsletters, press releases), held two conferences specifically to promote RDI as well as arranging for the project to be presented at other conferences, delivered two workshops aimed at key interest groups, and undertook a targeted contact programme with opinion formers.

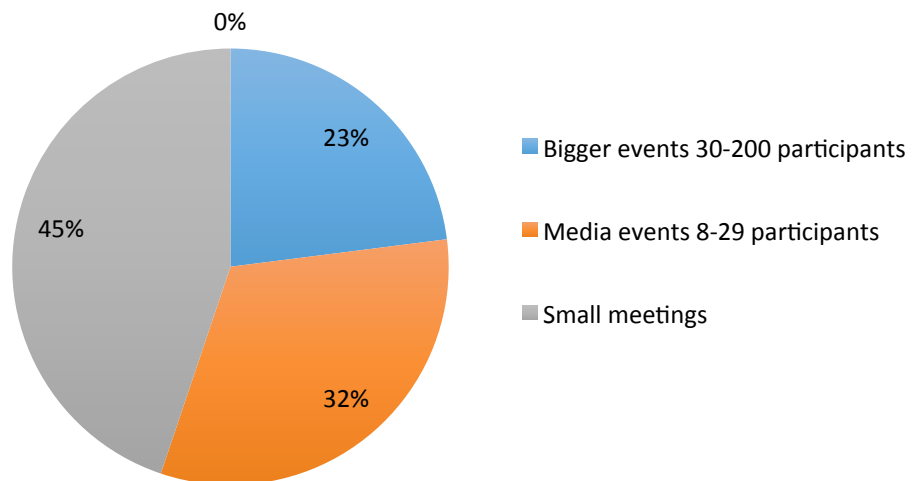
The two charts below illustrate the breakdown of the various audiences and at which types of meetings/events RDI has been promoted as well as the Linked Content Coalition:

⁸ <https://www.doi.org>

Audience covered



Type of events



There has also been significant exposure of the project beyond Europe at important events such as the presentation made to the **2015 EU China Round table on Copyright protection and development of Copyright** (Shanghai, China), and at the **DOI and IDF meetings that took place in Japan** in December 2015. The **USPTO** invited the Copyright Hub to present in April 2015 to a meeting on facilitating the on line licensing environment for copyright works in the US during which the LCC data model was presented. An important opportunity to showcase the project, together with the work and future objectives of the Linked Content

Coalition took place through a specific LCC and Copyright Hub presentation, followed by a Panel discussion at a high level conference organised by **WIPO** in April 2016, in Geneva⁹.

The dissemination activities will continue beyond the project in consultation with the Linked Content Coalition which is expected to take up the lead in moving forward the wider adoption of the LCC Framework.

An important learning from the various dissemination activities is that there remains a clear need for specific networking and dissemination initiative(s) covering several member states to ensure scale and reach, combined with a sector specific approach to provide information more generally about the LCC Framework and the RDI work to the various sectors (music, audiovisual, image, text), with a focus on how it can be implemented to benefit the creative sectors in the digital environment, and be a major enabler for the digital creative content market place as a whole.

3.5 RESULTS

RDI has moved the **LCC Framework** from being a paper specification to being a set of technical specifications and tools which can be made generally available to implementers.

Rights data of any kind can quite easily be **standardised, integrated** and then **queried** to meet a wide variety of needs, including discovering rightsholders and acquiring/granting licenses. The potential impact of widespread adoption of the Framework on access and automation in the digital supply chain is great, especially at the lower-price, higher volume end of the network.

The challenge of securing widespread adoption of the Framework is significant. Further technical development is needed to create production versions of the RDI prototypes, and potential providers need to be encouraged to create Framework-based services by understanding the benefits: this promotional role needs to be taken forward by the LCC, the Copyright Hub and others.

There are other factors which should help adoption. Because it is an interoperability model, the Framework does not significantly disrupt existing sector supply chains which in many cases operate satisfactorily, nor does it require data sources to make major changes in their own systems. Creating interoperable data should become an automated process at marginal operational cost which is more than compensated for by resulting benefits.

4 RECOMMENDATIONS AND FUTURE PLANS

The RDI project is finished and its consortium will be disbanded. During its lifetime, it was regarded as the prime mover in the practical development of the LCC Framework and tangentially of the Copyright Hub.

⁹ http://www.wipo.int/meetings/en/2016/global_digital_conference.html

4.1 RECOMMENDATIONS

A number of detailed recommendations relating to the future development of the LCC Framework and use certain RDI tools were made on the final report to the EC. These recommendations are summarised below. Discussions are ongoing between RDI partners, LCC and others to confirm whether (and if so, how) these recommendations may be implemented. As soon as some agreements are reached (hoped to be in July 2016) details will be included here and the final version of this report published.

Recommendation 1: The task of co-ordinating the implementation and adoption of the LCC Framework should return to the LCC.

Recommendation 2: The LCC should adopt, formalise, maintain and make available the RDI CRF, Ontology and mapping outputs.

Recommendation 3: LCC should co-ordinate a target implementation work aimed at establishing the Digital Rightsholder Statement (the "digital © notice" mentioned in section 3.1.2.1). This programme should include implementing a number of services which will be self-financing.

Recommendation 4: That the EC consider what political or financial support may be appropriate to support the LCC in expediting the implementation and wider adoption of the LCC Framework.

4.2 FUTURE PLANS

Each of the partners involved in the RDI technical work intend to build on or enable others to build on, what has been demonstrated.

Rightscom's work on the underlying model created by the Linked Content Coalition (LCC) will be taken forward by the LCC. Discussions are also under way between Rightscom and the LCC for the transfer of the RDI Ontology and mapping services to come under LCC governance, with one immediate goal to make these services available to the Copyright Hub.

The Copyright Hub has now scoped the initial open source software release from its technical partner Digital Catapult (expected in April 2016). Building on its early pilot implementations, the Hub is pitching this initial technical offering on "retail" licensing of use permissions, targeted at the low-value, high-volume end of the supply chain. The first release will be quite limited in scope and will use one specific schema (ODRL) for representing license data. The reason for this is practical, taking one step at a time in building a complex infrastructure.

Cineca will exploit the results achieved in RDI by integrating the various micro-service elements it has developed into new products and services. These micro-services are a data integration hub, transformation services, discovery services and rights schema. Initial implementations will include the use of the schema in the Forward Project for audio visual orphan works and iMediaCities in the field of movie rights, and of the hub platform for integrating rights data in the university domain.

Cineca and Rightscom jointly will look to exploit the mapping transformation and repository development work developed under RDI.

mEDRA's work on the Digital Object Identifier (DOI) has fed into the standards work of the International DOI Foundation and will lead to further development of the Handle System which underpins DOI, and has identified the need for harmonisation and standardisation within the DOI community of how to express and describe "relationships and representations".

Ediser is using its development to providing its collective management organisation (CMO) with a basic Reproduction Rights Organisation (RRO) system to manage the licensing services offered alone or in partnership with third parties, like Copyright Clearance Center. The CMO will be able to offer a better service to the rightsholders it represents (i.e. mandating publishers) by automating the work currently done manually, to innovate the way the licensing of secondary rights are managed (until now virtually confined to photocopy), and to offer more content available for licensing for a broader range of uses, with flexible policies, reducing barriers for end users to lawfully reuse copyrighted content.

RightsDirect has published the source code for the RightsDirect exchange with an open-source license so that it can be used by future implementers of the standards. In this way, the learnings and technology produced in the project create benefits for the entire community of implementers of the standards used in the RDI project. A version of this is being used by RDI partners for demonstration purposes.

CEPIC is presently looking into different possibilities to utilize the CEPIC IMAGE FINDER (CIF) concept in a way that will not be financially onerous for the organization and which does not interfere with the business model of their members. The CIF could be used as a tool by the images industry in different ways: among them to identify authors and right holders using traditional searching methodologies together with Visual Image Search (VIS) technology provided by different partners. Diligent Search compliance on Orphan Works and support to publishers and other potential users trying to locate right holders in order to license specific images are applications of the CIF that will be developed in 2016 and 2017, as soon as CEPIC members approve their development.

1. CONTACT DETAILS

5.1 PROJECT CONTACT DETAILS

Project website: www.rdi-project.org

Project Coordinators: Andrew Farrow, Godfrey Rust - andrew.farrow@rightscom.com,
godfrey.rust@rightscom.com

Communication: Ann Becker - ann.becker@europe-analytica.com

5.2 PARTNER CONTACT DETAILS

<p>Rightscom Ltd</p> <p>Suite 1866, Kemp House 152 City Road London EC1V 2NX United Kingdom</p> <p>www.rightscom.com</p> <p>Contact person: Godfrey Rust godfrey.rust@rightscom.com</p> <p>Phone: +442085798655</p>	<p>International Federation of Reproductive Rights Organisations (IFRRO)</p> <p>Rue du Prince Royal, 87 1050 Brussels Belgium</p> <p>www.ifrro.org</p> <p>Contact person: Olav Stokkmo olav.stokkmo@ifrro.org</p> <p>Phone: +3222346260</p>
<p>Coordination of European Picture Agencies Stock, Press and Heritage (CEPIC)</p> <p>Rue de La Mare, 46 75020 Paris France</p> <p>www.cepic.org</p> <p>Contact person: Sylvie Fodor s.fodor@cepic.org</p> <p>Phone: +493088910160</p>	<p>mEDRA SRL</p> <p>Corso di Porta Romana, 108 20122 Milano Italy</p> <p>http://www.medra.org</p> <p>Contact person: Paola Mazzucchi paola.mazzucchi@aie.it</p> <p>Phone: +390289280803</p>
<p>Consolidated Independent Ltd</p> <p>250 City Road, City Forum London EC1V 2QQ United Kingdom</p> <p>http://www.ci-support.com/public/</p> <p>Contact person: Kieron Faller kieron@ci-info.com</p> <p>Phone: +442077298493</p>	<p>RightsDirect BV</p> <p>Hoogoorddreef, 9 1101 BA Amsterdam Zuidoost The Netherlands</p> <p>www.copyright.com</p> <p>Contact person: Skott Klebe sklebe@copyright.com</p> <p>Phone: +19786462750</p>
<p>Consorzio Interuniversitario Cineca</p> <p>Via Magnanelli, 6/3 40033 Casalecchio di Reno BO Italy</p> <p>http://www.cineca.it</p> <p>Contact person: Gabriella Scipione g.scipione@cineca.it</p> <p>Phone: +390516171634</p>	<p>Europe Analytica Ltd</p> <p>Wiggens Green Helions Humpstead Jacobs, Haverhill CB97AD United Kingdom</p> <p>www.europe-analytica.com</p> <p>Contact person: Angela Mills Wade angela.mills-wade@europe-analytica.com</p> <p>Phone: +441865310732</p>
<p>Producentforeningen – Danish Producers’ Association</p>	<p>Album Archivo fotográfico SL</p> <p>Calle Juan de Austria, 126</p>

<p>Bernhard Bangs Allé, 25 2000 Frederiksberg Denmark www.pro-f.dk/content/danish-producers-association Contact person: Amalie Lyngbo Hjort al@pro-f.dk Phone: +4533862885</p>	<p>08018 Barcelona Spain www.album-online.com Contact person: Sergi Griño sergi@album-online.com Phone: +34932313740</p>
<p>Ediser SRL Corso di Porta Romana, 108 20122 Milano Italy Contact person: Paola Mazzucchi paola.mazzucchi@aie.it Phone: +390289280800</p>	<p>Getty Images Ltd 101 Bayham Street London NW1 0AG United Kingdom www.gettyimages.co.uk Contact person: Jonathan Lockwood jonathan.lockwood@gettyimages.com Phone: + 442032272900</p>
<p>European Writers Council Rue d'Arlon, 75-77 1040 Brussels Belgium www.europeanwriterscouncil.eu Contact person : Myriam Diocaretz m.diocaretz@inter.nl.net Phone: +3225510893</p>	<p>age fotostock Spain SL Calle Buenaventura Munoz, 16 08018 Barcelona Spain www.agefotostock.com Contact person: Alfonso Gutierrez alfonso@agefotostock.com Phone: +34933002552</p>