

# LCC and the “Digital Identifier Network”



Godfrey Rust, Rightscom/Linked Content Coalition

Keynote for IFRRO Business Models Forum, March 2013

## **Context: the digital content explosion**

The number of new and adapted works of all media types (text, image, audio, audiovisual) now loaded or created on the internet *each day* is greater than the total recorded analogue output of civilisation.

This figure was negligible ten years ago. The result is an unimaginable number of new daily **digital orphans** – content whose identity and rights are inaccessible to users or service providers in any automatable way.

15 years ago Amazon and Google didn't exist and they now dominate global markets. Yet we are just getting going: in the next 15 years, as internet use becomes ubiquitous, the rate of change and growth will be faster.

## **This presentation...**

Summary of LCC and the RDI project

Description of the **Digital Identifier Network** which is the whole context of rights management for the future (using the LCC's rights model to describe it)

What you do about it is up to you!

## Linked Content Coalition (LCC)

LCC was established in 2012 to develop building blocks for the expression and management of **rights and licensing across all content and media types.**

Membership global (includes IFRRO): all media types and all parts of the digital content supply chain.

Supported by EC (funding RDI project) and UK "Copyright Works" report leading to Copyright Hub.

Phase 1 just completed – "LCC Framework" this month.

Phase 2 in planning.



## **LCC Vision**

Use technology to benefit media supply chain participants,  
not to their detriment.



## **LCC Objective**

To be a catalyst to encourage the automated management of content rights in the digital network.



## Two LCC assumptions

An efficient rights data supply chain is pre-requisite for the efficient delivery of content to users and value to supply chain participants.

Rights data management is broadly the same in all media – differences of emphasis, not of fundamentals.



## **LCC is not...**

...advocating automation where it isn't appropriate

...about replacing existing standards like ONIX or ISBN

...biased to any sector or business model (including free use)





## Potential benefits

Easier discovery of rights ownership will **increase market size for rightsholders** and **decrease infringement**.

Increasing automation will reduce cost and **increase profitability** for all supply chain participants.

More standardisation will **lower system development costs**, encouraging **transformative innovation** and **increasing market size** for all supply chain participants.

Positively addressing perceived inefficiencies in the supply chain will **counter pressure from regulators** and **copyright law**.



## **First deliverable - LCC Framework (March 2013)**

Specification for best practice and interoperability in the digital rights data supply chain.

Specs for **Identifiers** and **Messages** (and, in future, **user interface/iconography**) in the digital network.

**Rights Reference Model** (RRM) – comprehensive data model for all types of rights in all types of content for all types of use and control.

RRM can be used for **system/message design** or data transformation for interoperability between other schemas (whether standard or proprietary) – a **“hub” model** to allow anyone to talk to anyone about rights.

## **RDI (“Rights Data Integration”) project**

Beginning May 2013, for two years, EC funded.

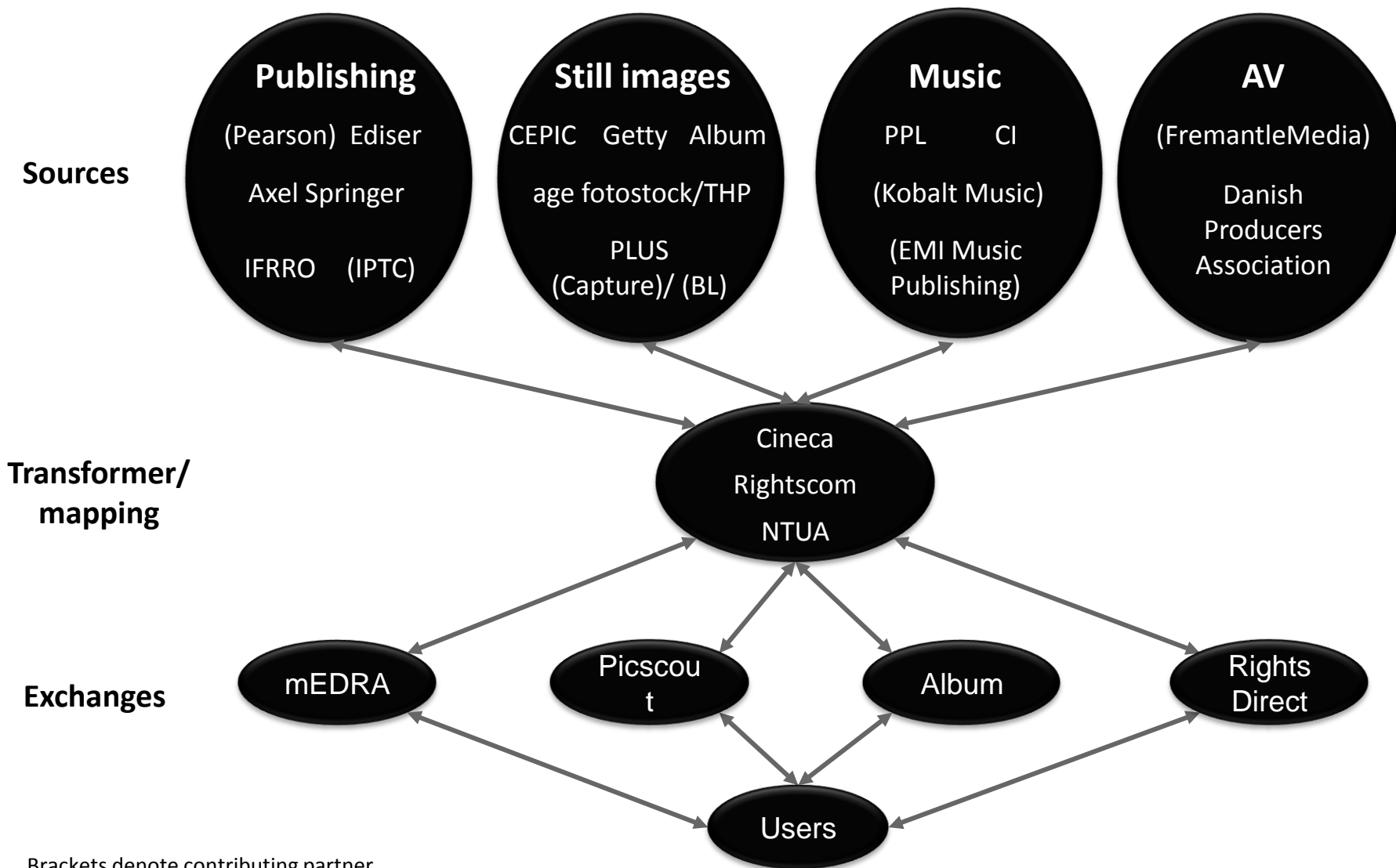
An exemplary implementation of the RRM as a “hub” (a prototype “Copyright Hub”).

Not “academic”: real businesses dealing with real data and hoping for real long-term business opportunities from RDI.

A range of data flows across the supply chain to show that a range of rights expression languages (standard or custom) from all media types can be transform and integrated using an implementation of the RRM.

Show how new standards can be implemented to fill gaps.

# RDI participants



Brackets denote contributing partner which is not a member of the consortium  
 Godfrey Rust IFRRO March 2013

## **UK Copyright Hub**

LCC endorsed by Hooper “Copyright Works Report”.

Hub working groups to review LCC Rights Reference Model with a view to recommending it as the data architecture for the Hub.

## **The Digital Identifier Network (“DIN”)**

The rest of this talk will use the RRM to present a vision of the present and future of the rights supply chain as a network of digitally resolvable identifiers.

What role a person or organization plays in the management of those identifiers will largely determine their usefulness in the digital future.

*Note: this presentation deals with the flow of rights from owner to user: it does not explicitly model the flow of usage information or money back from user to owner – but that follows the same route in reverse.*

## **The LCC Rights Reference Model (RRM)**

A comprehensive data model for all media and right types. Based on much previous work and best data modelling practise.

Designed to cover scope of all known existing rights standards and more. Extensible, flexible, optimizable.

Tested with use cases.

Data modelling is not rocket science: it is about describing the reality you want your system to deal with. A model should make sense to anyone who has some understanding of that reality. If it doesn't, your system will cause you trouble.

## **The RRM demystifies rights data management**

All multi-media rights data, however complex, can be aggregated and expressed in common and relatively simple ways.

Across-the-board interoperability is achievable - "our domain is special" is not true.



## Three distinctives of the RRM

A single "Right" Entity ("A state in which a Party is entitled to do something in relation to a creation, as a consequence of a law, agreement or policy"). Other models break rights into pieces without a common identity and treat them as attributes of content or licences: the RRM sees content and agreements as attributes of a Right.

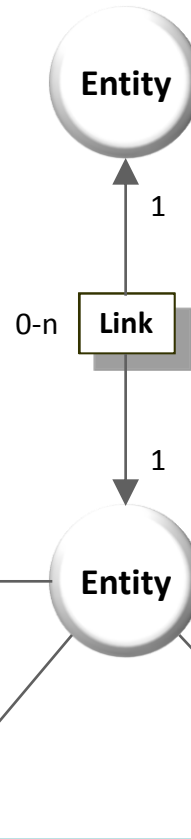
The full RRM covers the whole range of events and states needed for managing rights in a single model (including ownership, licences, assertions and conflicts).

The **LCC Entity Model** – an extensible way of modelling all Entities with the same five building blocks.

# LCC Entity Model

This diagram shows the common structure for each Entity in the RRM (and other models which LCC may specify in future).

Each Entity is built in a modular way from combinations of five types of Attribute, each of which has a different “micro-model” structure, exemplified here. Each Attribute is an Entity in its own right and may have Attributes of its own.



**Link**

A typed relationship between two Entities

*example*

<b>LinkType</b>	lcc:Creation_Party
<b>Entity1</b>	A123 (=“Moby Dick”)
<b>Entity1Role</b>	
<b>Entity2</b>	B987 (=“Herman Melville”)
<b>Entity2Role</b>	xyz:Author

**Category**

A categorization of of an Entity with a fully controlled data value

*example*

<b>Type</b>	xyz:RightType
<b>Value</b>	xyz:Play

**Time**

A point or period of time associated with an Entity

*example*

<b>Type</b>	xyz:ValidPeriod
<b>Mode</b>	lcc:Period
<b>Proximity</b>	lcc:Exactly
<b>From</b>	2012-01-01
<b>Proximity</b>	lcc:NotAfter
<b>To</b>	2013-12-31

**Descriptor**

A Name, Identifier or Annotation of an Entity in the form of an uncontrolled or partially controlled data value

*example*

<b>Type</b>	lcc:Name
<b>Value</b>	“John Smith”
<b>NameType</b>	ReferenceName
<b>Designation</b>	“Smith, John”, Indexed
<b>Part</b>	“John”, NamesBeforeKeyName
<b>Part</b>	“Smith”, KeyName

**Quantity**

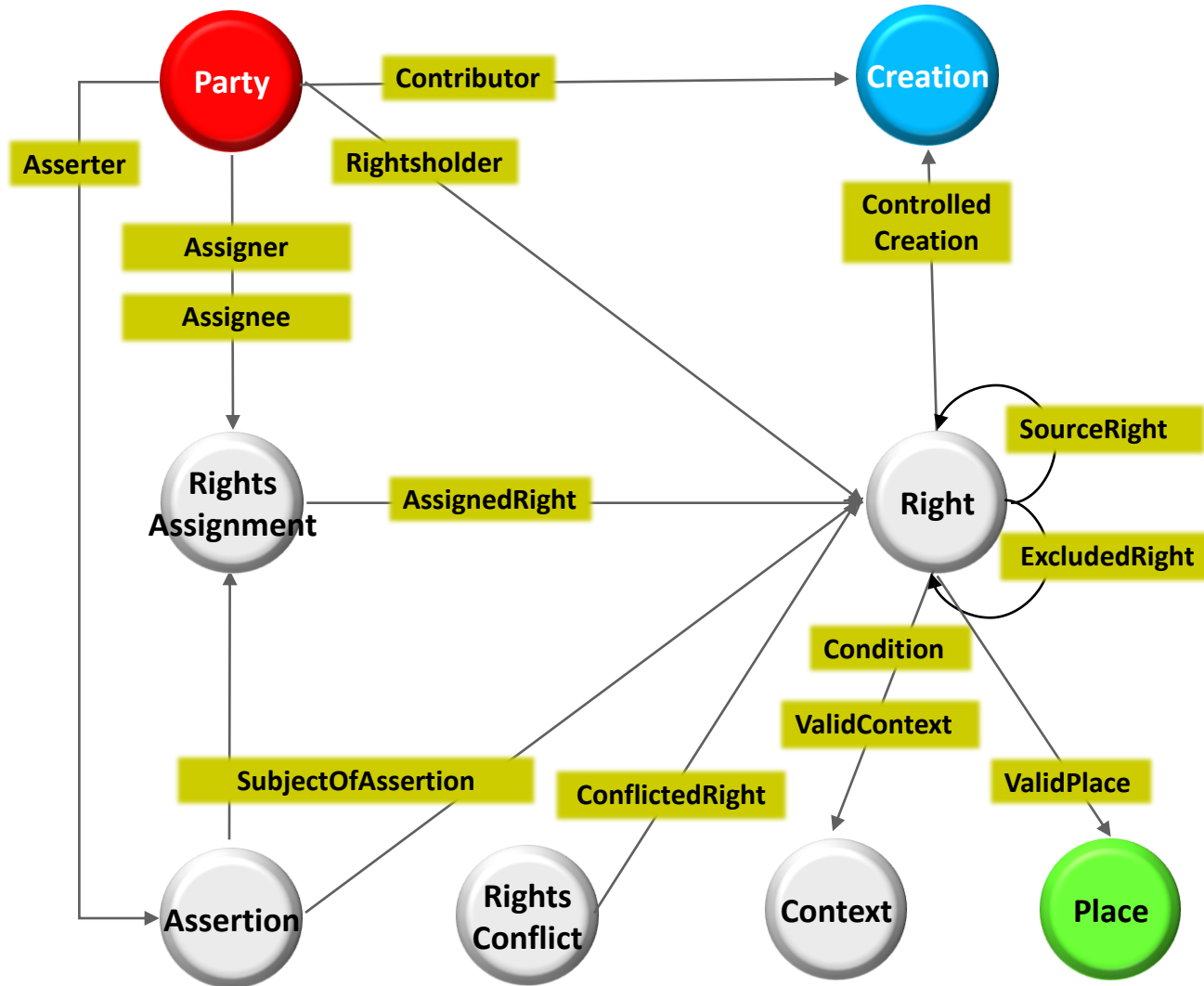
A measure of some aspect of an Entity

*example*

<b>Type</b>	xyz:FileSize
<b>Mode</b>	lcc:SingleQuantity
<b>Proximity</b>	lcc:NotMoreThan
<b>Value</b>	10
<b>Unit</b>	xyz:MB

# RRM Entities and principal Links

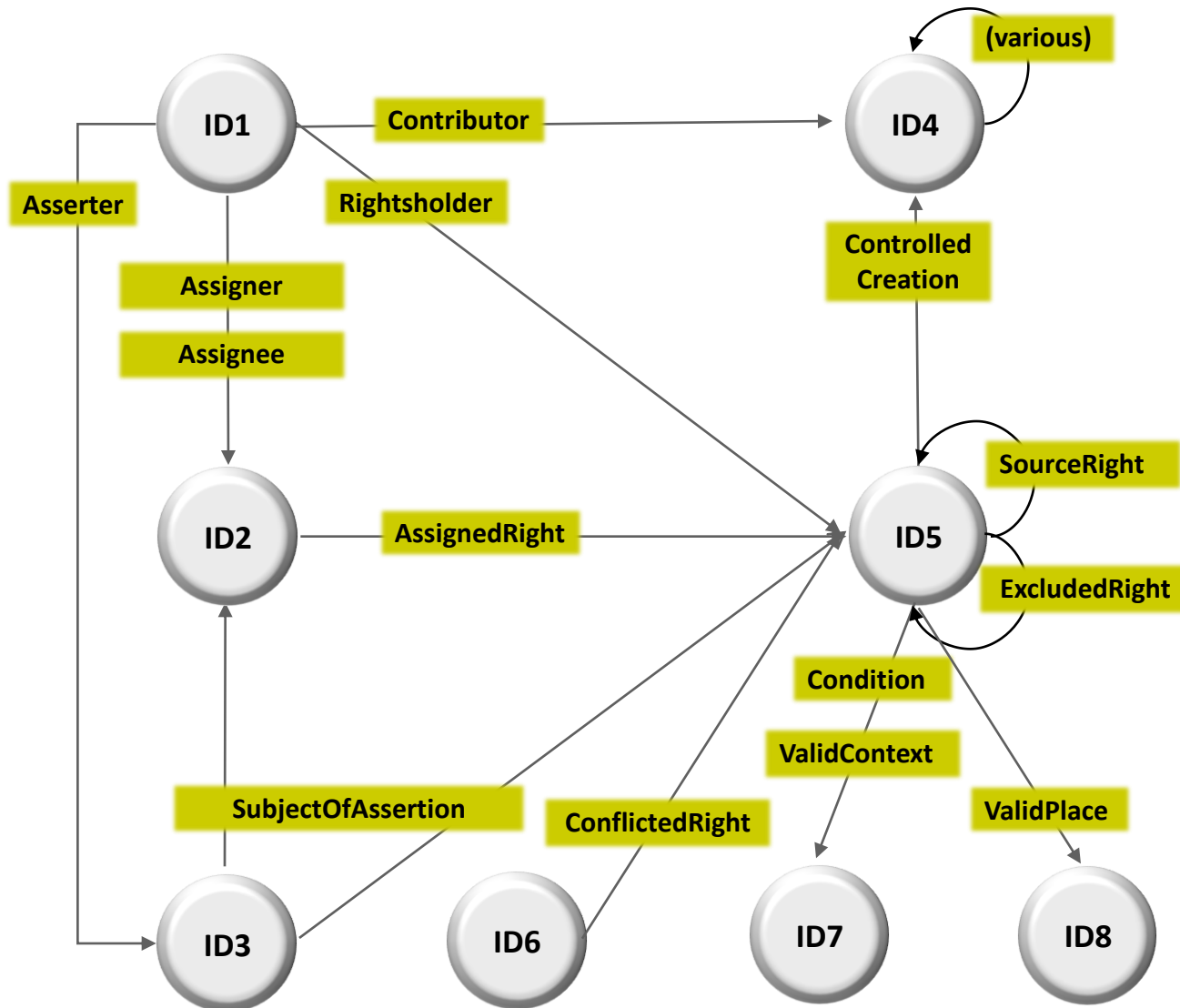
What the reality looks like



Each of the Entities is an Identifier

# RRM Entities and principal Links

What this looks like to a computer



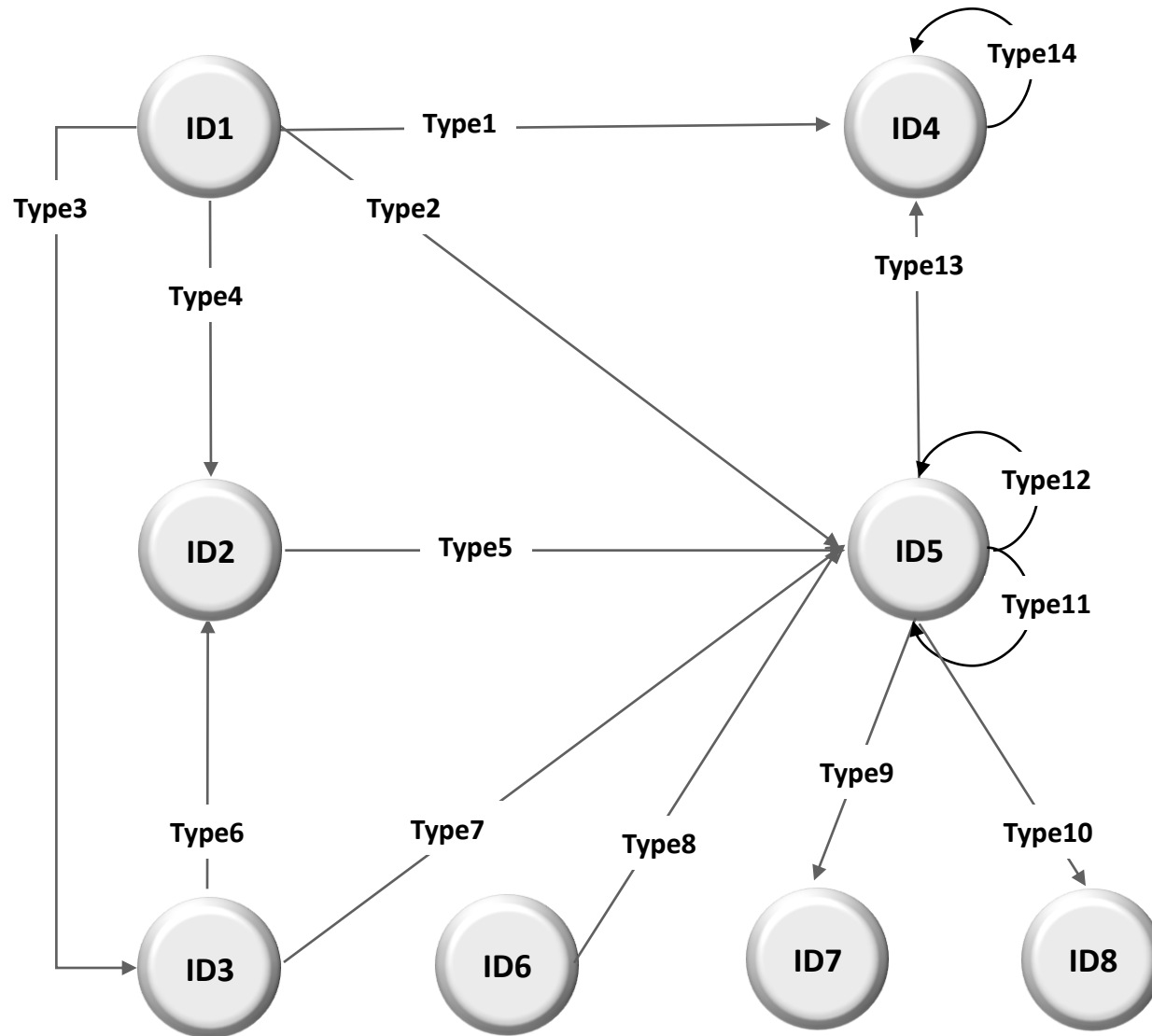
# RRM Entities and principal Links

What this looks like to a computer

Each of the Entities is an Identifier and each of the Links is defined by a Type (a term from a controlled vocabulary).

The computer has no idea what it means. It only works if IDs and Types are in a form that it can process.

It can recognise if an ID or a Type is the same as another – that is how it is able to pass information from one model to another.



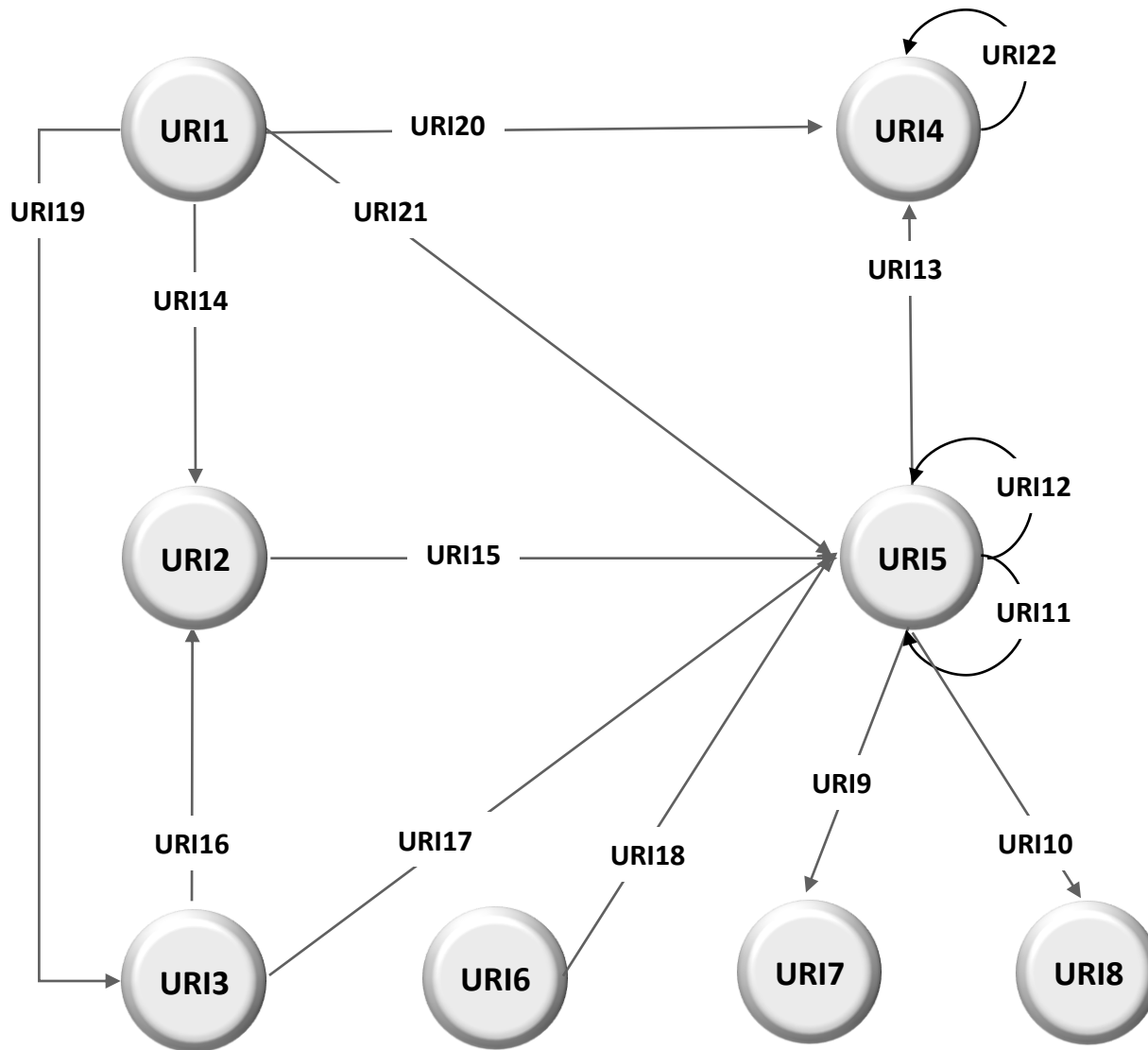
# The Linked Identifier Network

What this looks like (ideally) on the Web

On the Web, each ID and Type is best represented as a URI. That enables data be processed using standard Web protocols to make whatever connections are required.

This is what is creating the **Linked Identifier Network** on which digital commerce relies.

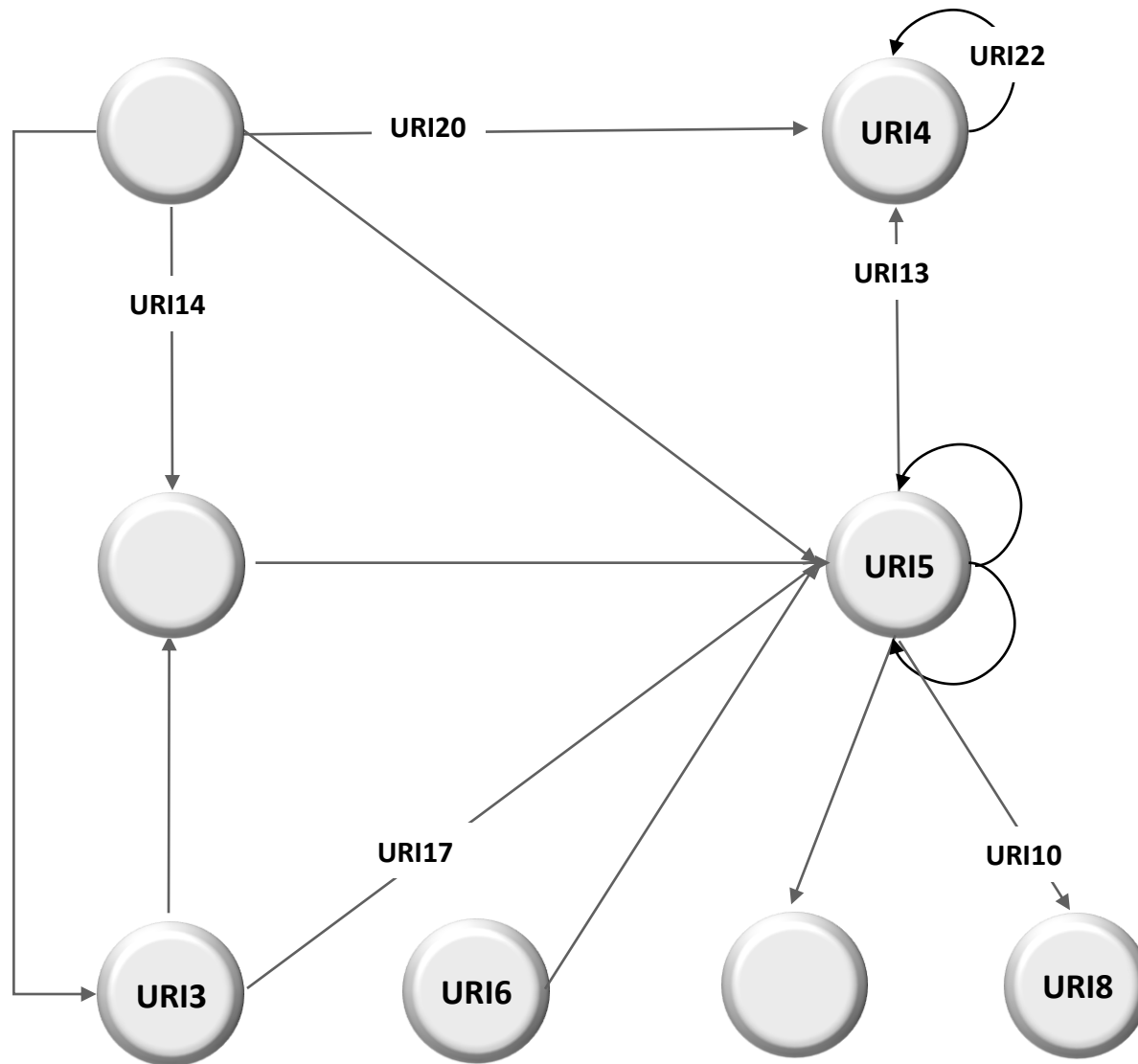
Rights, money, content, searches, piracy and anti-piracy actions all flow along this network.



Where there are gaps in the network, the Network is broken.

# The Linked Identifier Network

What this looks like (actually) on the Web



## **“The Network’s down”** (or not yet up)

The Linked Identifier Network is seriously broken (or rather, not yet built) in many places, and some of these get in the way of good content and rights management.

Either

- IDs and Types are missing, or
- IDs and Types are not mapped to each other where they represent the same things, or
- they are not yet usable as URIs

That sums up what LCC wants to see done.



## **Provisos**

Not all links are public. The network will always be confidential or inaccessible in some places for good reasons.

Not all parts of the network are worth the cost of building and maintaining them.

# **The future of the Digital Identifier Network**

Five things I expect to happen, and which LCC will be trying to facilitate:

## **The future DIN: Party IDs – first in the chain**

Party IDs are the starting point – and the blind spot in several sectors.

eg CISAC's great success story – the CAE/IP number.

eg ISNI (and IDs which link to it)

When Creations or Rights are being identified, the Rightsholders must first be identified if licensing is to be automated.

All Rightsholders who want them will have IDs (like ISNI, or an ID mapped to ISNI) which enable them to register content and rights. They will be able to get them quite easily.

## The future DIN: Creation IDs - “Digital declaration”

Automated content identification is already common in images (eg PicScout and YouTube) and audio.

Growing use in text world (eg Digimarc Guardian).

Allows a party to declare a **reference copy** – so a copy supplied to (eg) YouTube or PicScout allows service providers to offer discovery or monitor pirate usage.

At present this process is ad hoc and driven by specific business requirements (such as take-down).

Digital declaration will become the key to the authoritative “**once for all**” **declaration of works and associated Rights** throughout the digital network.

## The future DIN: Rights IDs

Rights data needs to be shared automatically much more than it currently is in the supply chain (volume, multi-media, re-purposing explosion).

Policies and Rights will have IDs issued at the point at which they are to enable resolution.

eg PLUS Coalition

Much but not all of this will be proprietary “behind closed doors”.

Assertions (“who says”?) and Conflicts will become critically important in managing distributed Rights data.

## **The future DIN: Controlled vocabularies (“Category IDs”)**

Already well established and widely used, these will be mapped through services like the Vocabulary Mapping Framework (VMF) and schema.org.

Apps will access Category IDs automatically and transform them seamlessly for different domains as required.

## **The future DIN: Resolvable IDs**

Creation and License/Rights IDs will allow people and systems to resolve automatically to different services – such as content, rights metadata or licensing systems, eg DOI/Handle already offers “multiple resolution”.

IDs will be embedded with content or in web pages, but metadata increasingly stored remotely – especially for rights.

## **“Declaration” vs “registration”**

This presentation talks about “declaration” rather than “registration” of digital content, to avoid a common misunderstanding.

“Registration” has several meanings, and is sometimes used to describe an act which establishes the existence of copyright in a creation.

“Declaration” here simply means providing a definitive, accessible copy of a creation along with metadata for the purposes of authoritative identification regardless of its copyright status (for example, “declaration” may also apply to works in the public domain).



## **The future of content and rights declaration (1)**

A person or organization with content and rights will be able to upload a Creation onto a digital service and:

- identify themselves and their roles (with IDs)
- identify their Creation(s) and links (with IDs)
- identify the rights they control and are willing to grant (with IDs)
- this process may apply to self-publishers or major corporations
- this process may apply to single works or complete repertoires

## **The future of content and rights declaration (2)**

The IDs they use may be standard or proprietary but will be “shared” with the network.

Controlled vocabularies from different namespaces will be interoperable.

Services will use apps and APIs so that for the user the process can be more or less automatic, using a set of default preferences.

With the RRM underpinning the standards it can be as simple or complex (and extensible) as it needs to be.

## **The future of content and rights declaration (3)**

Similar “ID at the point of activity” processes will support the issuing of licenses and usage reporting and monitoring.

There are no technical barriers to this – all the individual processes involved are commonplace and “big data” is no longer an obstacle.

The digital explosion described at the beginning makes this both necessary and inevitable.

# **Who manages the Digital Identifier Network?**

Like internet, all the participants.

Threat or opportunity for IFRRO members?

New services and intermediaries will appear, some of them growing very rapidly.

Your legacy data and systems can be an asset or a liability.

## **LCC – the next step**

LCC proposes to become a consortium of standards groups (eg Editeur, DDEX, IPTC, CC, PLUS Coalition, W3C, DOI, other ISO identifier agencies).

Enough of these already engaged with LCC to secure continuity.

Anyone else (eg IFRRO) may be affiliated and participate in, propose and/or fund projects.

**Project to facilitate “digital declaration at the point of activity” will be top of the agenda.**

Thank you for your time and attention.

godfrey.rust@rightscom.com

[www.linkedcontentcoalition.org](http://www.linkedcontentcoalition.org)