The last ISO SC9 ad hoc ID interoperability meeting agreed to await the outcome of the Linked Content Coalition work as the basis of a way forward, so that is the main item for our meeting this year. With the agreement of the SC9 secretariat, Godfrey Rust, the technical lead of LCC, will attend the ISO SC9 ad hoc ID interoperability group to discuss the LCC work.

The LCC main documentation has now been published (http://www.linkedcontentcoalition.org/#documents/cvjv). Of particular relevance to the Identifiers registration authorities are:

- The Digital Identifier Network, a narrative background paper*
- The LCC Principles of Identification v1.0

A useful introduction to the concept of the LCC is in the first section of the introductory document, The LCC Framework v1.0, which states: “The principal goal of the LCC 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.… The next steps are the first implementations of the LCC Framework”.

The LCC Principles of Identification paper provides eight recommendations as a model of best practise for identifiers in supporting the highest level of automation, trust and accuracy within the supply chain and network. Detailed support for the recommendations is provided in an appendix (comprising both a review of key principles and a survey of major existing identifier systems). The recommendations from this paper (but not the lengthy appendix) are attached as part of the current document.

In addition to the technical papers which are output from the LCC, a further paper is in preparation on the LCC continuation and its future operations. It is to be accompanied by a covering letter and a two page note which informally describes a number of actual (RDI) and exemplary projects which might be undertaken through LCC. This will be distributed to all those standards and industry bodies associated and working with LCC (e.g. ONIX, DDEX) and also to the wider SC9 group of RAs. It is hoped that this will have been done by our SC9 meeting, and so this can also be discussed.

It is open to the Ad Hoc group to determine if and how the principles outlined and followed by LCC should be adopted. Note that the LCC principles:
• have been adopted already by the Copyright Hub:  
• will underpin the forthcoming EU-coordinated RDIO (Rights Data Integration) project
• and of course are already implemented in the earlier activities which LCC builds on (ONIX  
  and DDEX messaging schemes, etc.)

The issue of the endorsement of the technical outputs of LCC is a separate (though related)  
matter to the possible endorsement of any plan for LCC’s future.

*Note: as of May 3, *The Digital Identifier Network* narrative background paper has not been  
published on the LCC web site. It is hoped that this will be available in some form before the SC9  
June meeting.
Principles of identification

Version 1.0, April 2013
Editor Norman Paskin, Identifier Workstream Leader, n.paskin@tertius.ltd.uk

This specification comprises the LCC recommendations for the design and use of identifiers within the digital network in the digital rights supply chain. Detailed support for the recommendations is provided in the attached appendix.

The eight recommendations here are presented as a model of best practise for supporting the highest level of automation, trust and accuracy within the supply chain and network. They are not "mandatory" in the sense that none of them is legally or systematically enforceable for all identifier types, and failure to comply will not normally block the supply chain entirely, only make it more time-consuming, labour-intensive and error-prone.

1. Identification is essential
   - Each entity which needs to be recognised distinctly in the digital network should have at least one public or shared identifier.
   - Types of entities here include those identified in the LCC Rights Reference Model: Party, Creation, Place, Context, Right, RightsAssignment, Assertion, RightsConflict.
   - In particular, this includes each item of content which needs to be recognised (at whatever level of granularity is required), and each person or organization who is recognised as (or claims to be) a contributor or rightsholder of content (an "interested party").
   - A public identifier is not necessarily humanly readable: "public" means that it is accessible to people or machines within the digital network.

2. Identifiers should not contain dynamic or confusing "intelligence"
   - In general, "dumb" identifiers (that is, identifiers whose characters or elements have no intended meaning or referent) are preferable as they avoid some obvious pitfalls, but a limited "intelligence" can be safe and useful, and on occasion essential.
   - Information about the type of the identifier is normally safe and useful: for example, prefixing an ISBN with "ISBN".
   - Information about the issuer and date of issue of the identifier is best kept out of the identifier itself if possible in human-readable identifiers of content, as it is easily and commonly misinterpreted to refer to the owner or publisher of the content and its date of creation or publication. However, many identifier standards incorporate one or both of these references so they are often a fait accompli, and so the onus is on the parties or systems using them not to make false inferences.
   - Persistent information about the entity (that is, information that should not change) should not be encoded within the identifier, because (a) like all metadata,
it may be interpreted differently in different contexts and (b) it may be found to be incorrect at a later date. All such information should be declared as metadata, to which the identifier may resolve. Some existing identifier standards (such as V-ISAN) do encode or imply metadata about the referenced entity (for example, that it is of a certain type or has certain properties) and so again this must simply be managed as well as possible.

- **Dynamic information** about the entity (that is, time-limited metadata such as status codes or rights ownership) should never be encoded in an identifier.

- These principle apply to digital fingerprints or binary identifiers created automatically from the digital structure of an item (for example, for recognising specific images or audio tracks) as well as identifiers created independently. Digital fingerprints do of course encode information about the entity they identify, but these are not-human readable and provided the content they identify is not itself dynamic this "intelligence" is normally benign and of course of enormous value for content recognition.

### 3. Identifiers should be resolvable

- A **resolvable** identifier in the digital network is one that enables a system to locate the identified resource, or some information about it (such as metadata or a service related to it), elsewhere in the network.

- Because the World Wide Web is the dominant network using the Internet, then it is a minimum requirement to support the Web, and a potential requirement to support other networks. This in effect recognises the URI as a primary practical common framework for global digital content identification. Non-URI identifiers may still be used where appropriate but should be expressible as or within URIs where necessary.

- The URI syntax can incorporate existing standard or proprietary identifiers while remaining globally unique, and much technology already exists for recognising and resolving URIs in various ways. Resolution is essential, and on their own many existing ID standards (being pre-digital in origin, such as ISBN, ISRC and ISWC) don’t natively support this but require a URI “prefix”.

### 4. Identifiers should be capable of multiple resolution

- An identifier should be capable of being resolved to more than one location for different types or instances of metadata: for example, to find least one basic description and one statement of rights.

- Choices in multiple resolution may be made by human beings or by machines following rules.

- Multiple resolution should be capable of managed change as data sources change: flexible resolution is essential to allow legacy and proprietary systems to interact.

- Multiple resolution of an identifier should be possible without special knowledge except for the ability to communicate using standard technical protocols.

- Multiple resolution requires a basic and extensible standard “typing” vocabulary of resolution so that different services (in the example given, different metadata
types) can be automatically located. This approach is common and usually implicit within proprietary closed systems but is not yet generally recognised as an inevitable requirement of open linked data.

5. **Identifiers should be accessible**
   - Content identifiers should be accessible to users by (e.g.) embedding them where possible within the item of content or its message sidecar during interchange; making relevant information available in metadata; or embedding identifiers on webpages to support resolution to various services; and so on.
   - Different approaches are useful for different purposes; the aim should be to provide accessible persistent identification.

6. **Identifier registration should be under well-defined registry operations and policies.**
   - “Linked Data” technologies alone are not sufficient to establish a trustworthy industry-standard data exchange. The identifier-registered material must be ‘data worth linking to’: curated, value-added, data, which is managed, corrected, updated and consistently maintained by registration authorities and agencies. The LCC specifications should enable "curated data", i.e. consistent, managed, linking enabling links to other "quality data" with confidence, while still capable of using existing Linked Data technologies.
   - Adequate supporting descriptive and rights metadata should be declared along with a registered identifier to support discovery and avoid ambiguity. Metadata should be registered under some method of governance (a registry or registration procedure) to ensures its authority and its ongoing maintenance in locations to which the identifier may resolve, using defined service types. Metadata about an entity is commonly declared by more than one party, and registry procedures can therefore provide ways of identifying the asserters of particular items of metadata and facilitate the resolution of conflicts within different metadata declarations.
   - **Trust** in the accuracy and persistence of identifiers and their supporting authoritative metadata is critical. Accountability for persistence can only be ensured through a governed registry arrangement, where there are also provisions for maintaining metadata after the original assessor is defunct, dead or otherwise unwilling to accept responsibility. This does not necessarily mean a central repository of metadata, but it requires a registration procedure supporting identification. Mechanisms are needed to minimise instances of several Parties issuing identifiers for the same content, where creation or original publication is shared. Mechanisms are also required for dealing with duplication (the issue of more than one identifier to an entity) and ambiguity (the issue of the same identifier to two or more different entities).
   - Trust is necessary for several steps in identification:
     - that the identifier is for the thing you believe is being identified;
     - that the resolver you are using is the resolver you think you are talking to;
that the resolver you think you are talking to is actually the right resolver for the job;

- that the data in the resolver relates to the thing you are asking about;
- that the data in the resolver has been put there by a party with authority to do so;
- that data in the resolver hasn't been subverted since it was registered.

7. **Metadata associated with an identified Entity should be published in standard form**

- Metadata associated with an identified entity should be published in extensible and interoperable syntactic formats (such as RDF, JSON or XML) using formalised schemas with defined elements and using controlled vocabularies wherever possible. The specification and definitions of the schemas and vocabularies should be freely available to those needing to interpret the metadata.

- Standards may be public, formal, de facto or proprietary: there will always be a diverse range of metadata schemas for different sectors and functions (and this trend is likely to continue to increase).

- The semantic mapping of any well-formed schema to another (as described in the LCC specifications for metadata interoperability) cannot compensate for poorly-defined or ambiguous source data.

8. **The asserter of Rights metadata should be identified**

- Authoritative rights metadata associated with an identifier should be formally “asserted” so that its provenance is clear. The asserter is not necessarily the same Party as the provider or publisher of the metadata or the rightsholder: for example, an intermediary such as a collecting society, agent or licensee may create or pass on metadata on behalf of a party further up the supply chain, such as a creator or publisher: it is the party on whose behalf the metadata is declared who is the asserter or authority. An intermediary may therefore legitimately and necessarily publish conflicting metadata from different asseters on occasions, especially about rights ownership. It is a requirement for any metadata aggregator to have policies and methods for managing conflicting data from different sources, and on occasions from the same source.

The following are related recommendations for follow up by the LCC successor organization:

- The Vocabulary Mapping Framework (VMF) should be used for mapping metadata (terms and schemas).
• The Vocabulary Mapping Framework (VMF) should be mirrored and/or expanded from its current coverage to cover needs for similar mappings for other entities, and attention be given to active maintenance and a governance structure of VMF.
• Develop a scheme and methodology for associating a given existing non-internet registry scheme with a URI and associated structured metadata.