

Submission of Paul Jessop, Founder and Director, County Analytics Ltd

**In response to
Library of Congress - US Copyright Office
Notice of Inquiry Docket 2014-1
“Strategic Plan for Recordation of Documents”**

Date: March 14th, 2014

Introduction

Paul Jessop, Founder and Director of County Analytics Ltd respectfully submits these comments in response to the Notice of Inquiry (NOI) issued by the Copyright Office on January 10, 2014 concerning the Copyright Office’s *Strategic Plan for Recordation of Documents*.

Paul Jessop is a consultant providing services in fields including metadata and media identifiers. His clients include organizations that own, represent or manage rights in content in the US and worldwide. He has wide experience in international standardization and sits on the relevant standards committees for media identifiers in the US, the UK and France. He also participates in the committee of the International Organization for Standardization (ISO) that develops standards in this area. He is the Head of the UK Delegation to the parent committee that oversees this work along with numerous library-related standards, and has previously acted as Head of the UK Delegation to the Moving Pictures Expert Group (MPEG).

He was previously Chief Technology Officer at the International Federation of the Phonographic Industry (IFPI) in London and subsequently at the Recording Industry Association of America (RIAA) in Washington DC. In these capacities he supervised numerous projects concerned with rights management and rights exploitation in the recording industry. He was one of the project managers of the Music Industry Integrated Identifiers Project (known as MI3P) which led to the foundation of Digital Data Exchange (DDEX) and continues to represent the recording industry associations at meetings of that organization.

He manages the recording industry’s involvement in the international standardization aspects of the International Standard Recording Code (ISRC) and acts as Executive Director of the US ISRC Agency on behalf of RIAA. He also represents the interests of CISAC, the International Confederation of Societies of Authors and Composers at the registration authorities of the identifiers for musical works (ISWC), audiovisual works (ISAN), textual works (ISTC) and parties (ISNI).

These comments are submitted solely as a personal response to the NOI. Their preparation has not been supported by any client of County Analytics Ltd and they do not represent the views of those clients.

Background

Just as the Copyright Office notes that the present notice builds upon comments received in response to the earlier notice on *Technological Upgrades to Registration and Recordation Functions*, this response builds on the response the submitter made to that earlier notice. The previous response is posted on the Copyright Office website¹.

This response focuses on the issues raised in:

- Question 1 on a guided remitter model,
- Question 2 on structured electronic documents,
- Question 3 on linking documents records and registration records and
- Question 4 on the use of standards for identifiers and other metadata.

This response is silent on the issues raised in question 5 whose implications are outside the expertise of the submitter.

Overall Recommendation

The Copyright Office should recognize that those who interact with its systems vary greatly in their technical sophistication and the flexibility they have to adapt their information technology infrastructure to accommodate external interfaces. This is not related directly to size: some larger users may have lengthy development cycles while some smaller users can react quickly. On the other hand small users may not have the resources to make the required changes.

Further, the Copyright Office should note that users who seldom record documents will find it hard to justify the cost of developing systems that connect the data used in commerce to processes used to communicate with the Copyright Office.

The Copyright Office should adopt a strategy that starts with "quick wins", especially those that impose no costs on users, and build on these to develop a comprehensive system over time.

Specifically, the Copyright Office should adopt a robust but flexible abstract data model that allows the various entities it deals with (and the relationships between them) to be identified and described in a standard way. This would include not only the copyrightable works that are to be registered and whose transfers are to be recorded, but also the rights in them and the parties registering and recording them, their representatives and any services that are used to convey this information.

¹ URL: http://www.copyright.gov/docs/technical_upgrades/comments/County-Analytics.pdf

Such an abstract data model would be applicable to both manual and automated processes.

The data model should be compatible with those used in industries whose works are the subject of copyright and those used by librarians and archivists.

The data model should also recognize that economics encourage registration of collections of works or portions of works that are used separately in commerce. For example a whole journal issue may contain several articles and the articles may be further split into sections that are licensed separately. The elements that are licensed may not be determined until after registration so the data model may need to reference data sources where identification of the elements is defined. This will likely be outside the Copyright Office's systems but interoperability with these systems is important, as is discussed below..

The Copyright Office should ensure that all its interfaces with users (registrants, enquirers, researchers, librarians, litigants, other government departments and foreign administrations) migrate over time towards using this data model however the interface is structured – manual or automated.

The implications of this recommendation are set out in the following sections.

Interoperability with Databases Outside the Copyright Office

It was noted in the earlier response that interoperability with databases operated privately would add to the value of the information held by the Copyright Office. It is suggested here that such interoperability might in the long run both (i) reduce the burden placed on users (whether registrants or inquirers) of Copyright Office services and (ii) reduce the Copyright Office's operating costs (and hence user fees) that result.

For example, a sound recording may be identified by its "metadata" – for instance its title, recording artist and recording (or release) date (in practice some further information may be needed). In commerce it will likely also be identified by an international standard identifier (in this case an International Standard Recording Code – ISRC). This identifier will be recorded in systems that manage the sound recording within the industry and can be used in communication with the Copyright Office. If the relationship between the ISRC of the recording and the Copyright Office's record number for a copyright registration (the SR number) is stored in the registration database, it does not very much matter which is cited in communication though there some issues that need careful consideration. As noted above, a registration may cover a group of recordings and the same recording may appear as a track on several registered works (i.e., albums). Additionally several related recordings (with different ISRCs) may rely on the same copyright registration. These might include the original unedited version, an edited version which can be sold without a "parental advisory" notice, a radio remix and a ring tone.

Where this interoperability is created, several benefits might accrue. Firstly as noted below external data could be used to assist in the guided remitter process so that accuracy is enhanced. Perhaps more importantly, those seeking information on recorded documents relating to a copyrighted work will have a greater chance of finding the right documents.

A Guided Remitter Responsibility Model of Electronic Recordation

The notice intimates that there are relatively few events (assignment, licensing etc.) that can trigger the need for a document to be recorded. There are also relatively few types of work (textual work, sound recording etc.) that can be implicated by such an event. This means that the basic options of a guided model will be able to cover the vast majority of relevant events.

The notice notes that predictive models can be used to focus the energy of the checking process on those areas where errors are most likely to be found. This response goes further: the guided remitter process could make effective (though selective) use of external information sources to guide the predictions. Though the information in an external database cannot be seen as authoritative for copyright registration purposes, it can be good enough to prompt for obvious errors (such as the swapping of two digits in an identifier). Further, this may lead to the ability to correct errors in external databases (to the benefit of the whole sector) if the prompt indicating a possible error is overridden.

Some of these databases are community curated and freely available. For others, the source of corrections may mean that a mutually advantageous arrangement for access can be negotiated.

The system must of course be designed to ensure that deliberately falsified submissions can be deterred, detected and corrected. Equally, accidental errors must be capable of correction after they are detected.

It is important that the data model adopted by the Copyright Office is compatible with (i.e. maps cleanly to) the data models used by external databases. Otherwise the effort involved in making the external data usable in the predictive process may exceed the benefit obtained.

The volume of database records at the Copyright Office containing the relevant identifiers will initially be limited. However over time this will likely improve.

Submission of Recorded Documents

As the documents to be recorded must accompany the recordation request and the intention is to move away from a requirement for a physical signature, the Copyright Office should ensure that the supported formats used for the image of the

recorded document are published and open. Permitting the use of proprietary formats would seriously degrade the value of the documents recorded to future generations if the decoding technology is not available.

Structured Electronic Documents

The adoption of a standard data model as recommended above allows the specification of formats for the provision of information to and from the Copyright Office.

The aspect of this that must be stressed is the difference between (i) the structure of the data, (ii) the data itself and (iii) the presentation of the data. To take a simple example, a car registration number may have a structure of three letters followed by three numbers. A particular car may have the registration number "ABC123". However for different purposes this can be presented in different ways: in correspondence from the government department it appears as laser printed characters but as applied to the car, the characters are embossed on a metal plate. In databases accessible to law enforcement and others, the registration number appears as a binary encoded data block. We accept this as a routine way of dealing with data.

In a similar way, a notice that the rights in a work have been assigned has a structure: the identity of the assignor, the identity of the assignee and the names of the works assigned. The data itself is straightforward for a particular assignment but the presentation of this data can take many forms: a paper document, a computer image of a document, a database record or an electronic message. Importantly, translating between these formats is straightforward also, at least in some directions (going from a paper document to an electronic message is harder).

Well-established technologies such as XSLT² enable this kind of conversion. This means that if users are encouraged to make submissions using (say) an XML message, they can be converted (without data loss) into formats that are human readable. They can also be ingested into databases or transformed into further messages for other purposes.

The Copyright Office should not underestimate the effort required from users to transition from current manual processes to automated ones. As noted above, if there is a low level of documents submitted for recordation, the change may not in fact be justified. The Copyright Office should recognize this and make these more efficient processes optional unless there is a consensus that they can be supported without material inconvenience to users. However, the documentation of specifications in widely understood formats (such as formal schemas) will

² <http://en.wikipedia.org/wiki/XSLT>

encourage the provision of communication tools by third parties (who will themselves see a large enough market to justify the development effort).

Linking Document Records and Registration Records

The approach outlined above will create the technical capability to include information specifying the copyright registration. The question of whether this linkage should be required – for instance by mandating the provision of registration numbers for all registered works that are the subject of a recorded document – is different and raises policy issues beyond the scope of this response.

The Use of Standards for Identifiers and Other Metadata

The earlier response from County Analytics Ltd (and several of the other responses to that Notice) made a strong recommendation for the use of standard identifiers and metadata formats in copyright registration. This would maximize the interoperability between Copyright Office systems and external databases, and hence increase the utility of the data stored by the Copyright Office to users. The same arguments apply to the recordation process.

For a recorded document to be useful to someone seeking to understand the current and past ownership of the rights in a work, they must be able to find any such recorded documents that apply to the work in question and that implies that the document must in some way be associated with identifying information about the works covered. While textual descriptive information can go some way to deliver this, the use of codified identifiers is much more robust, provided the association between the work and its identifier is itself authoritatively established somewhere. That said, not all works will have assigned identifiers and the assigned identifier will sometimes not be available to someone wanting to use it.

The association between identifier and work is critical and is typically made by including “reference metadata” in a registry alongside the identifying code. This reference metadata is sufficient descriptive metadata to distinguish (“disambiguate”) a work from apparently similar works.

For some classes of work, the definition of reference metadata may be difficult. For photographs, there are often many very similar shots in a set and using metadata alone may not be useful. In these cases, consideration may be given to using a registry that makes a link direct to the content. This may involve storing a digital version of the work itself (such as a thumbnail image of a photograph) or a digital fingerprint of the work.

Whatever technology is used to create the association between work and identifier, the authority and stability of the registry is critical and the Copyright Office should ask searching questions about these issues before deciding to use any particular identifier and its registry in Copyright Office operations. Some registries are, as

noted in the Notice, privately run, others are run under the auspices of an international standards setting organization such as ISO, while others are community-curated. Each brings some benefits but also potential hazards.

Whether the use of such identifiers should be made compulsory in recorded documents, or whether incentives should be created to encourage remitters of documents to be recorded to use them is again a policy issue beyond the scope of this response. It may be noted that many such identifiers are in regular and routine use in the commerce that depends on copyright for its existence. However, the participants in this commerce may not yet have integrated the systems that drive commerce with those that drive copyright legalities and will not prioritize this development unless there is a strong business case for it. In the long run, users who adopt such identifiers to communicate with business partners might also be able to do so in communication with the Copyright Office but this will likely not happen quickly. This supports the recommendation above that the early publication of an abstract data model will allow this development to take place in a phased way over time without wasted effort.

Conclusions

There is considerable scope for the Copyright Office's systems to be updated to allow a move towards the use of standard identifiers and standard messaging systems to make those systems more useful to all stakeholders. Interoperability with other systems would allow more trustworthy searching and aggregation of information about works that have been registered.

The submitter is very willing to provide more information on these recommendations on request and to participate in further stages of this important process.