

Benjamin K. SemelDirect Tel: 212-326-0131
bsemel@pryorcashman.com

July 13, 2020

Via email

Regan Smith
General Counsel and Associate Register of Copyrights
Jason Sloan
Assistant General Counsel
United States Copyright Office
Library of Congress
101 Independence Ave. SE
Washington, DC 20559-6000

**Re: Docket No. 2020-5
Music Modernization Act Reports of Usage – Metadata**

Dear Ms. Smith and Mr. Sloan,

We represent the Mechanical Licensing Collective (the “MLC”) and write in response to your June 30, 2020 letter requesting responses to questions from the Copyright Office (the “Office”) concerning whether digital music providers (“DMPs”) should be permitted to report modified metadata to the MLC or instead be required to report metadata in the original unmodified form as acquired.

The MLC responds to the Office’s questions directed to the MLC as follows:

- 1. The DLC explains that “[i]t is not uncommon for DMPs to fill in empty data fields when the relevant information is known to them,” and that track title alterations made by at least one of its members are “in every case at the request of the labels.” Please clarify the MLC’s desired treatment of data in the following scenarios:**
 - a. where data is acquired from a source other than the sound recording copyright owner or other distributor (e.g., the DMP adds data to an otherwise blank field);**

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 2

- b. where data is modified by the DMP at the request of the sound recording copyright owner or other distributor; and**
- c. where data is modified through some automated process initiated by the sound recording copyright owner or other distributor, such as an updated data feed.**

Should the data in any of these scenarios be considered “modified” or “altered” data for the MLC’s purposes? If so, why?

a. Metadata acquired from a third-party source that is not the sound recording copyright owner (or its authorized representative) should not be substituted for metadata from the sound recording licensor in the monthly reporting. However, where a sound recording copyright owner has failed to provide any metadata for a particular field, and the DMP has filled the blank field with metadata received from another source, the MLC would not consider such metadata to be “altered” so as to trigger a preclusion on reporting. As the MLC has noted, its focus is on ensuring that source metadata is reported, rather than precluding reporting of additional metadata. Thus, the MLC does not seek to preclude DMPs from reporting metadata added to fields left empty by sound recording copyright owners. Of course, DMPs must continue to undertake the necessary efforts outlined in the regulations to obtain the respective metadata from the sound recording copyright owners, at which point that source metadata should be reported.

b. Whether metadata changed by a DMP “at the request of the sound recording copyright owner or other distributor” should be considered altered metadata for reporting purposes depends on whether the “request” is direction from the sound recording copyright owner or distributor to update and correct the metadata in accordance with its records, or whether it is a modification made based upon DMP demands or requirements. If, for example, a sound recording copyright owner conveyed generally to DMPs a request to update Title metadata for a particular licensed sound recording, the new title should qualify as metadata “acquired from” the sound recording copyright owner. If, on the other hand, a sound recording copyright owner “requests” that a DMP alter Title metadata of licensed sound recordings to fit the DMP’s style, the resulting new titles should not be considered metadata acquired from the sound recording copyright owner.

c. The critical question here is not who initiates the modification or whether the modification is automated or unautomated. The determinative question is whether the modified information is, in fact, metadata coming from the sound recording copyright owner, and whether it is provided to make a correction in accordance with its records, rather than to meet a demand for alteration by a DMP. Thus, if a DMP demands that a sound recording copyright owner use an automated process to modify its metadata to conform to the DMP’s particular style or specifications, the MLC would consider the resulting metadata altered. On the other hand, the

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 3

MLC considers metadata “acquired from” sound recording copyright owners to include such metadata *as altered* by the respective sound recording copyright owners, including through regular data feeds provided to DMPs. Metadata acquired from sound recording copyright owners is not merely the first transmission of metadata, but also subsequent updates (including alterations) to that metadata from the sound recording copyright owner.

- 2. For each data field proposed in § 210.27(e)(1) that the MLC understands DMPs to revise, re-title, or otherwise edit or modify in the ordinary course, describe how and to what extent receiving reported data in altered form would materially impede the MLC’s matching efforts. The MLC has explained that differences in the number of characters in a title can be relevant to an automated string comparison algorithm. Please respond to the DLC’s assertions that “any reasonably sophisticated matching algorithm would be able to handle th[e] alterations [made by DMPs].” For example, if the title string comparison algorithm does not yield a match, what other matching strategies will be employed? If the MLC believes that the reporting of altered data in the fields at issue would frustrate those additional matching efforts, please explain, and include illustrative examples.**

As a threshold matter, insofar as DMPs alter metadata “in the ordinary course,” such “ordinary course” is for marketing and not royalty administration purposes. But the statutory purpose for which the DMPs are to report metadata received from sound recording copyright owners is to allow the MLC to identify sound recordings and match them to musical works. 17 U.S.C. 115(d)(4)(A)(ii)(I)(aa). It is thus critical for DMPs to be required report the sound recording copyright owner metadata, because such approach is what the MLC needs to fulfill its matching and ultimately its royalty administration and payment functions. To be clear, DMPs may continue to have the metadata altered to create versions for their marketing purposes, but would just also preserve the version from the source.

On the question of “how and to what extent” altered metadata will impede the MLC’s matching process, the algorithmic process that underlies the matching platform is not susceptible to linear or binary paraphrasing. The matching process is not a question of looking behind Door #1, and if nothing is found, looking behind Door #2, and so on, until the answer is found behind one door. Multiple fields are simultaneously compared, and the result is not a binary Yes or No, but a confidence level based on the multifactor comparison (and even that is a reduction of a sophisticated, multilayered process). The DLC’s inexcusably oversimplified statement that the MLC should “be able to handle” a piece of altered metadata glosses over the issue. While a matching algorithm may not be fully defeated by a minor or cosmetic change to a single metadata field, the alteration of metadata makes the algorithms harder to maintain, and reduces the

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 4

confidence levels and thus the automated matching rate regardless of how sophisticated the algorithms are. Plain and simply, altering metadata will reduce the effectiveness of any metadata matching system. The argument that a single piece of altered data can be “handled” misses the point entirely.

Furthermore, altered metadata is not just an automated matching problem. Uses that do not match automatically (a subset that would increase with the reporting of altered metadata) need to be matched manually, whether by MLC staff or by copyright owners through the MLC portal. Both of these jobs are made more difficult by having to sort through the multiplication of use record variations caused by DMP alterations to metadata.

The MLC firmly believes that nothing in the DLC’s statements on this issue rises to reasonable grounds for allowing the reporting of intentionally altered metadata, rather than the source metadata that meets the statutory purpose for reporting, i.e., identification and matching. The MLC further notes that every other commenter that has weighed in on this issue has underscored the serious problem with allowing reporting of altered metadata. (*See, e.g.*, Sony & RIAA Ex Parte Letter (Dec. 9, 2019), at 2; A2IM & RIAA Reply Comments, at 2-3; Paul Jessup Initial Comments at 2-3.)

- 3. To what extent, in time, expense, and/or increased amount of unmatched works, does the MLC estimate that permitting reporting of data in certain fields altered for DMP display purposes will affect its matching efforts? In your answer, please consider that the proposed rule would require the reporting of many fields unaltered, including the ISRC and other universal identifiers, and the DLC has represented that over 99% of tracks will be reported with an unaltered ISRC. Please include any other information that would be helpful for the Office to consider.**

It is simply impossible to quantify the time, expense and/or increased amount of unmatched works that will be caused by unknown amounts of alteration of metadata by an unknown number of DMPs with an unknown level of variation between the different alterations. The MLC cannot quantify that impact now (nor can the DLC with reductive arguments).

The DLC’s claim that this is a minor issue because the DLC alleges that over 99% of tracks have unaltered ISRCs is empty and fundamentally misleading. To begin with, the DLC has only yet connected with a handful of DMPs in the marketplace—it has demonstrated no idea about what the metadata looks like for the other DMPs, and the MLC has to match the uses of every DMP.

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 5

Moreover, as the MLC has repeatedly stated, ISRC codes do not suffice for matching.¹ The MLC could receive from fifty DMPs, in a worst case for a single sound recording, fifty variations on its title, fifty variations on its artist, and fifty variations on its version title, along with other data inconsistencies.

Moreover, although the “one percent of tracks” figure is neither accurate nor relevant, even if it were, one percent of tracks is a *staggering* number. The number of tracks at issue is not the size of a typical DMP’s massive library, but the size of all DMP libraries combined, since the MLC must match each track as reported by each DMP. One percent of Apple Music’s library is over 600,000 tracks. Add the libraries of just Spotify, Amazon, Google/YouTube, Pandora, Tidal, Deezer, SoundCloud, Qobuz and MediaNet, and the “one percent of tracks” looks to translate to over 6 million tracks² and growing every day. And this is only a fraction of the total combined libraries of the full marketplace. If altered metadata obstructs automated matching on even a small fraction of 1 percent, it could amount to millions of works thrown into manual matching, which could amount to literally hundreds of human work years reestablishing matches.

These variables highlight the dramatic asymmetry between the lifts required by the two sides of this issue. Each DMP by definition can receive and possess the unaltered metadata from its sound recording licensors (and if the DMP had any of that metadata altered already, it could just have the unaltered metadata resent by the licensors, as is routinely done for a variety of reasons). The task of transmitting that unaltered metadata to the MLC is simple and straightforward. And while retaining and providing unaltered metadata may require some DMPs to make a one-time workflow change, once they make that change—separating out unaltered source metadata needed for administration purposes and altered metadata used for marketing purposes—both interests will thereafter be served, and the process will operate automatically with no friction. The MLC will have the metadata it needs and that the statute requires, and the DMPs may continue to further modify their separated, altered metadata for their marketing purposes. The rest of the DMP’s practices around altering metadata for its display and style purposes need not change at all. In contrast, for the MLC to receive only altered metadata is an ongoing and constant drain and wear on its systems, making its automated processes harder to maintain and less

¹ As the MLC explained in its February 26, 2020 *ex parte* letter, there is no single, global database that matches all ISRC codes with other metadata fields, and there are incorrect ISRC codes in use, including through reuse and piracy, with no authoritative voice that can take a DMP’s ISRC listing and identify with certainty what sound recording, let alone musical work, was actually streamed to the user. ISRC codes are just one of multiple metadata fields that must be used to identify sound recordings and match them to musical works.

² https://en.wikipedia.org/wiki/Comparison_of_on-demand_music_streaming_services

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 6

effective, and also compounding the amount of manual review required, increasing costs and decreasing efficiency.

One of the core purposes of the MMA is to improve identification, matching, and royalty processing and payment. Forcing the MLC to use the same altered metadata that the DMPs used that contributed to the system that the MLC was created to fix is inconsistent with the statutory goals.

4. Please describe the MLC's approach to matching efforts with respect to other causes of data inconsistencies that the MLC will need to address. Can the MLC describe the anticipated additional level of effort resulting from reporting of data in certain fields altered for DMP display purposes, as compared to these other data inconsistencies?

There is a fundamental difference between metadata inconsistencies that are the result of isolated errors, and those that are intentional and institutionalized. Errors in metadata will undoubtedly occur from time to time. But that is not a reason for permitting DMPs to intentionally provide inconsistent metadata.

Moreover, there is no "inefficiency cap" when it comes to metadata inconsistencies. Rather, each additional metadata inconsistency compounds the previous one and makes the process even harder as they synergise with each other. This is precisely why it is so important to preclude reporting of intentional alterations of metadata, which is very easy to avoid, as there will no doubt be other metadata inconsistencies (missing fields, incorrect ISRCs, ISWCs and other identifiers, metadata reported in the wrong fields, etc.) that will be unavoidable.

For the reasons noted above, the MLC cannot predict the precise import of the unknown extent of altered metadata problems across the entire DMP marketplace (including the large segment of the DMP marketplace that is not yet part of the DLC membership) as compared to the import of other metadata problems that will surface across the entire DMP marketplace. However, the MLC can confidently state that altered metadata will be a force for reducing matching efficiency and effectiveness, and will only compound the negative effects that arise from other metadata inconsistencies.

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 7

5. **Are there any non-statutorily-enumerated data fields proposed in § 210.27(e)(1) that the MLC either does not intend to use for matching or does not believe will lead to materially better matching results?**
 - a. **If so, please identify which fields.**
 - b. **If not, describe the usefulness of each of these fields.**
 - c. **In your response, please consider the following comments. ARM suggests that such fields “(e.g., distributor, album title, UPC, catalog number) are unlikely to be materially helpful for matching in most instances.” SoundExchange says: “[I]t does not seem important for the MLC to learn from service providers every possible detail about sound recordings and musical works—and particularly information about ownership and payment that is most subject to change and most likely to be inaccurate when reported by a service provider. What is important is that the MLC receive from service providers enough data points about sound recordings and musical works to be able to match reported usage to known repertoire, or to have clues about where to look for definitive information about previously unknown repertoire.” And in connection with a separate examination of fields to be reported by the MLC to copyright owners, Songwriters of North America (“SONA”) and Music Artists Coalition (“MAC”) state that “[a]lthough all of the nonmandatory categories are important, we wish to emphasize in particular UPC codes, which are sometimes the only reliable way to identify the particular product for which royalties are being paid.”**

All of the metadata fields proposed in §210.27(e)(1) will be used as part of the MLC’s matching efforts. Fields that do not currently feature in automated matching algorithms can be of particular value for the MLC’s manual matching efforts when automated matching does not succeed (for precisely that reason, if no other), as well as by copyright owners working through the portal to match their works. The cited comments of SONA and MAC are illustrative here, reinforcing that songwriters and copyright owners rely on a wide variety of metadata to match their works and help ensure that the proper people are paid.

Regan Smith and Jason Sloan
United States Copyright Office
July 13, 2020
Page 8

The MLC is available to provide further information or respond to further questions from the Office upon request. The MLC appreciates the Office's time and attention, and looks forward to participating in the upcoming telephone conference on this matter.

Sincerely yours,



Benjamin K. Semel