verifi media

MUSIC RIGHTS & DATA COLLABORATION

The Key to the Next Level of Revenue in the Music Industry

Bill Rosenblatt

September 2023

GiantSteps
Media Technology Strategies

Executive summary

The music industry is healthy again after recovering from the mortal wounds of the file-sharing era. 80% of global recorded music revenues now come from digital sources.[1]

And while traditional streaming revenue growth has started to slow down in certain geographies, the industry today is hitting several inflection points in music creation, marketing, distribution, and consumption. Now labels, music publishers, and service providers are all looking for additional ways to sustain revenue growth and profitability for the longer term.

One major source of additional growth today is the explosion in user-generated content on social platforms such as TikTok. This is leading to accelerated growth in synchronization (synch or sync) licensing, which is also growing due to the more active roles that music catalog owners are now taking in marketing music assets for new types of revenue. The key to all of these growth opportunities is improving management of music, ownership, and licensing data. Better data management leads to more complete, accurate, and timely royalty collection as well as increased opportunities for licensing.

Although the industry has made progress in recent years through adoption of standard identifiers, metadata schemes, and messaging protocols, many opportunities remain through data collaboration.

Data collaboration—when different entities communicate through services to match, complete, and maintain their data—enables significant incremental revenues and efficiencies in a range of use cases across the entire digital music supply chain. Data collaboration is already used in various other industries with demonstrated financial benefits.

In this white paper, we'll review the structural data problems that remain in the music industry despite the adoption of standards and messaging protocols. We'll discuss the economic consequences of poor data management for many of the participants in the supply chain, including music publishers, catalog investment funds, labels, CMOs, and DSPs.

Then we'll describe data collaboration and show how it builds on existing data management practices while avoiding the pitfalls of the single centralized database approach that was tried unsuccessfully in the previous decade.
We'll provide several examples of financial benefits from data collaboration in the music business that are hypothetical but based on real-world scenarios. From these examples,
we can estimate that data collaboration can lead to revenue increases in the region of 20% for music rightsholders.

Introduction and background

Streaming growth is slowing down, but music industry revenue growth can continue through the explosion of user-generated content.

The music industry has returned to growth after the decline of more than 50% in the US in the decade following the shutdown of the original Napster file-sharing network. Starting in 2015, worldwide music revenue has increased an average of 10% per year.

The industry is on track to exceed the inflation-adjusted revenue peak it achieved in the late 1970s, if not its all-time peak in 1999, within the next couple of years.

At the same time, revenue growth from streaming is slowing down in geographies such as the U.S., Europe, and Australasia, while synch licensing has taken over as the fastest-growing revenue category globally.[2] Still, the industry is expected to grow steadily over the next several years; Goldman Sachs projects more than 8% annual global revenue growth in recorded music revenue through 2030.[3]

Music consumption also continues to grow worldwide through increased adoption of streaming services; Luminate Data measured over 3.3 trillion audio and video song streams in the first half of 2023, a 32% increase over the first half of 2022.[4]

Revenue growth from streaming is continuing to accelerate in geographies such as Africa, the Middle East, and Latin America.
Various other digital-age innovations are contributing to further growth. One is the continued explosion of user-generated content on social platforms that use music, such as TikTok, which enable users to create short video clips that include portions of songs.

GLOBAL RECORDED MUSIC INDUSTRY REVENUES 1999 - 2022 (US$ BILLIONS)[5]

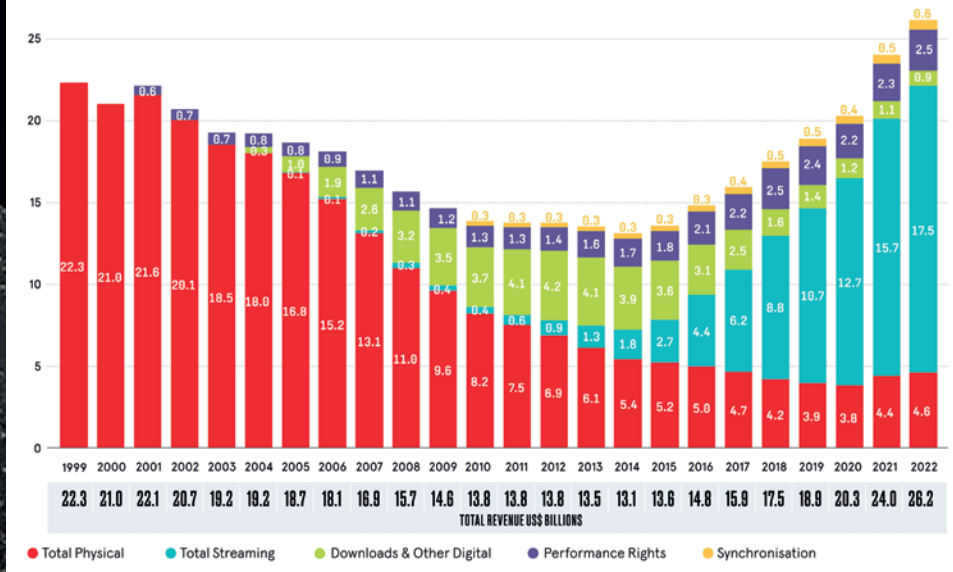


FIGURE 1

TikTok now boasts over a billion active users worldwide; more than 400 songs included in TikTok videos received more than a billion views in 2021,[6] and the major labels have made license deals with TikTok to allow for monetized use of their catalogs in users' videos. This increased use of music in videos is one of the main factors contributing to growth in synch revenues of more than 20% annually worldwide.

Other innovations contributing to further growth are new tools and services, from Ableton Live to VEVA Collect to Bandlab that are making it easier than ever for users to create new content from existing music and post it online. Remixes and mashups are on the rise. And emerging generative AI tools are expanding the possibilities of music creation and production on a larger scale than ever.

The opportunities for growth in music monetization are even greater amid all of these innovations. At this point, limitations in monetization opportunities come from limitations in data – the data about creators, rights, and royalties that flows among music industry entities worldwide. We'll see various examples of this later in this whitepaper.

The industry needs to go to the next level in maintaining and communicating structured data that is complete, accurate, and up to date to take advantage of these opportunities.The good news about music, compared to other forms of content (such as TV, film, and books), is that it deals in units of commerce that are well understood across the industry: music has songs (compositions) and tracks (sound recordings), which embody musical compositions and can be organized into collections (albums) which are released as products (LPs, CDs, etc.).

The bad news is that these units of commerce are involved in large numbers of transactions that have to take place every time someone streams a song, uses it in a video, or uses a sample in another song; those transactions involve multiple types of licenses which arise from the dual nature of copyrights in musical compositions and sound recordings, and they often involve multiple entities whose identities, attributes, or ownership shares may not be fully known or accurately represented.

Digital music services pay multiple streams of royalties, which derive either from copyright law or from industry convention:

Sound recording royalties for reproduction and distribution (known as master rights), paid to record labels or digital distributors such as TuneCore, CD Baby, or DistroKid.

Sound recording royalties for public performances (sometimes known as neighboring rights), paid to sound recording PROs (sometimes called neighboring rights organizations) or labels.

Composition royalties for reproduction and distribution (known as mechanicals), paid to publishers or CMOs/MROs.

Composition royalties for public performances, paid to composition PROs or publishers.

If videos are involved, synchronization (a/k/a synch or sync) licenses are usually required, meaning that royalties are paid on compositions and (if the original recording is used) sound recordings to the respective rightsholders.

Growth in music monetization is limited by data management practices.

Standard commercial practices for these units of commerce have developed over the history of the recorded music industry starting over 100 years ago.

Organizations have been set up over the years to administer and process transactions according to these rights: music publishers, record labels, distributors, collective management organizations (CMOs, PROs, MROs), publishing administrators, aggregators for independent artists, and so on.

Each of these organizations has historically managed its own data and done its own rights administration.

All of these organizations have been able to collaborate at some level thanks to the introduction of standard identifiers such as ISRC, ISWC, UPC, IPI, and ISNI (see [SIDEBAR 2]), which are managed by disparate entities.

The most commonly used identifiers in the music industry are these:

ISRC, *created in 1986 as a unique identifier for sound recordings, administered by IFPI.*

ISWC, *created in 1995 as a unique identifier for musical compositions, administered by CISAC.*

UPC, *and closely related* **EAN**, *identifiers for physical products (such as LPs and CDs), designed in the early 1970s to be used with barcodes, administered by the international supply chain standards organization GS1.*

IPI, *created in 2001 as a unique identifier for songwriters and music publishers, administered by CISAC and BIEM.*

ISNI, *a more recent standard launched in 2011 and increasingly used in the music industry as a unique identifier for recording artists, administered by the ISNI International Agency in the UK.*

There are also standards for communicating metadata related to rights administration. CISAC manages standards for communicating information about musical compositions among songwriters,
music publishers, and CMOs.

On the sound recording side, the global standards body DDEX manages a family of several metadata standards that are built into their standard messaging protocols. DDEX standards are most often used in communications from labels to DSPs and vice versa but are increasingly being used for other purposes, such as discovery metadata and communication of data to CMOs and other rights administrators.

Some of the most widely used CISAC standards for communication of composition information:

CWR, *used by music publishers to send information about compositions to CMOs. CWR files contain information about compositions including title and ISWC of the work as well as information about publishers, sub-publishers, acquirers (all with their IPI numbers), and their splits (ownership shares) and territories.*

CCID, *an older standard used by some CMOs to send information to DSPs about royalties they are claiming for use of compositions embodied in sound recordings.*

CRD, *used for communication of composition usage among CMOs and publishers.*

Some of the most widely used DDEX standards for communication of sound recording information:

ERN, *used by sound recording licensors (record labels, indie aggregators, etc.) to feed release information to digital music service providers (DSPs).*

DSR, *used by retailers and DSPs to report sales, usage, and/or revenue from music to rights holders or administrators.*

CDM, *used by some rightsholders and CMOs to send information to DSPs about royalties they are claiming for use of compositions embodied in sound recordings.*

Other DDEX standards relevant to rights and royalties administration include **BWARM** (Bulk communication of Work And Recording Metadata), **RDR-N** (Recording Data and Rights Notification), **MWN** (Musical Works Notification), and **RIN** (Recording Information Notification).

1 Structural Data Problems

Despite the growing adoption of these standards, the acceleration in growth and diversity of the music economy today is leading rightsholders to routinely leave money on the table; we'll describe specific examples of this below. The situation will only get worse as the industry's size and complexity continues to increase.

There are over 300 licensed music DSPs in the world; CISAC's 225 affiliated organizations collectively represent over 5 million creators;[7] as of the first quarter of 2023, DSPs were collectively adding 120,000 tracks to their catalogs every day worldwide, up from 93,000 in 2022;[8] and royalty transactions from streaming alone number in the billions daily. All these numbers will continue to grow substantially in the years to come.

At the heart of the problem is inadequate or insufficiently structured data: music data, metadata, and licensing data.

There are many reasons for this, starting with the classic "garbage in, garbage out" (GIGO) problem that exists in every field. Yet in the music industry there are various important sources of structural metadata deficiency; many of these came to be through industry conventions or historical precedents that are not fit for a digitally driven business.

We'll describe a few of them here.

Despite the adoption of standards, rightsholders routinely leave money on the table.

Lack of composition metadata at DSPs

As mentioned above, record labels and indie distributors deluge DSPs worldwide every day with new music tracks, accompanied by feeds of metadata, typically in DDEX ERN format. Yet these feeds usually do not contain complete composition metadata such as ISWCs, music publishers, and related splits; and they often contain songwriter information that is incomplete at best.

The lack of composition data in labels' feeds to DSPs dates back to the early days of streaming, when independent streaming services had to hire licensing agencies to go find compositions that matched the recordings that labels sent them, find the rightsholders for those compositions (and their splits), and pay the mechanical royalties.[9] Once interactive streaming became popular in the 2010s, the inevitable errors and omissions from this process became consequential enough to lead to lawsuits and searches for solutions.

Yet solutions to this problem have been slow in coming. One came in the United States the form of the Music Modernization Act of 2018 (MMA), which established a non-profit Mechanical Licensing Collective (MLC). The MLC acts as a centralized clearinghouse for matching plays of sound recordings on streaming music services to their underlying compositions and disbursing mechanical royalties to publishers and songwriters under statutory mechanical licenses in U.S. law. These rightsholders submit composition rights ownership information to it directly.

Feeds to DSPs often don't contain data on rights ownership for the compositions in the recordings.

Otherwise, a typical process for DSP royalty payments is as follows: DSPs send usage reports every month or quarter to CMOs. These include data on sound recordings used and types of usage (paid subscription stream, tethered download, etc.). Because DSPs

do not typically get complete composition information from labels and indie artists in a timely manner, they are unable to pass it on to CMOs in usage reports.

Therefore, CMOs often have to figure out composition rights owner information by themselves, a task that can involve parsing billions of items of play data to determine the compositions and their associated publishers, rightsholders, and possibly other CMOs that may be involved.

Some CMOs have databases of compositions and matching recordings which they can use to find this information, but even then, they may need to communicate with publishers and other CMOs—possibly in other countries—to get this information. This time-consuming process may still result in missing, obsolete, or inaccurate data.

Once a CMO has this information, it can send "claim files" to DSPs containing information about composition royalties that the DSPs owe; the DSPs then pay those royalties.

Yet the need to consult multiple sources and possibly issue corrections later can take time and delay payments all the way down the chain.

ROYALTY PAYMENT PROCESS FROM DSPs

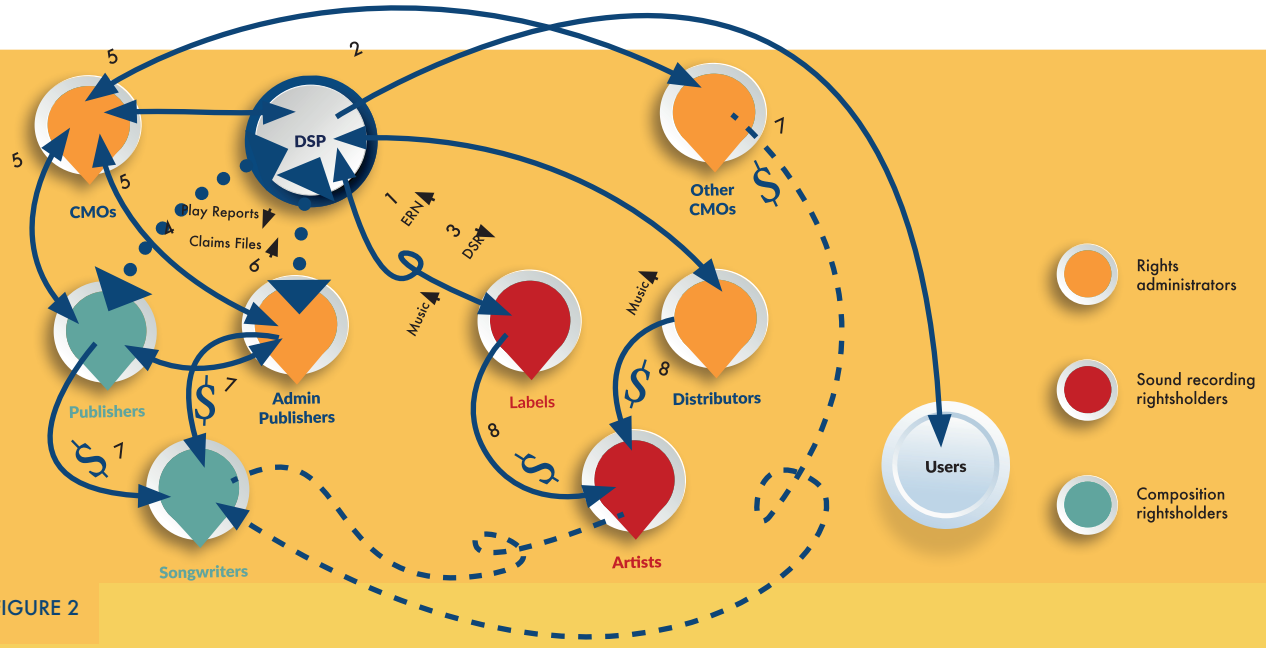


FIGURE 2

In a typical process for a DSP to pay royalties to music rightsholders, the steps are as follows, assuming that songwriters have all registered their works with their CMOs:

1 Labels and distributors send music along with metadata feeds in DDEX ERN files to a DSP.

2 Users stream music on the DSP.

3 The DSP sends DDEX DSR files to labels and distributors after the end of each reporting period (month or quarter). This data includes ISRCs to identify sound recordings, along with use counts, types of uses, and other information.

4 The DSP sends play reports to the local CMO for calculating composition royalties. It also sends play reports to publishers with which it has direct licenses.

5 The CMO determines the compositions embodied in the sound recordings and their songwriters, publishers, and splits. It communicates with publishers, other CMOs, and possibly third parties in cases where it can't figure this out.

6 The CMO sends the DSP a claim file, which could be in CISAC CCID or DDEX CDM format, for that reporting period showing which rightsholders are implicated in the music played and what royalties are owed.

7 The DSP pays royalties to the CMO, which passes them on to rightsholders or other CMOs as necessary, or it holds up the royalty payments because of issues with data in claim files.[10]

8 The DSP pays sound recording royalties to labels and distributors.

9 For example, overlapping claim files from multiple sources could claim more than 100% of rights in a given composition (known as "overclaiming").

DIGITAL MUSIC BUSINESS COMPLEXITY

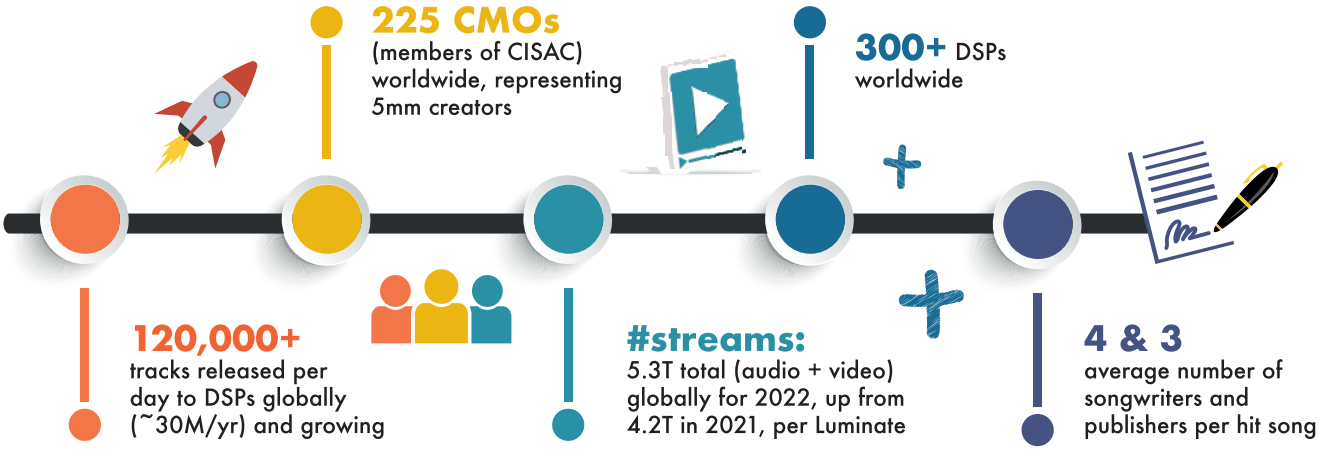


FIGURE 3

Many CMOs must match recordings to compositions and find complete splits data for compositions.

The figure on the left shows a typical case of composition rights data flow in a country that has a single "home CMO" that administers digital royalties from DSPs. This is just a simple example; many variations are possible.[11]

In other words, a typical process involves CMOs bearing the responsibility for two tasks: matching recordings to compositions, and finding complete rights ownership data (splits) for the territory in question. In some territories, the home CMO doesn't process digital royalties; in such cases, the DSPs may engage rights management services on their own behalf so that they can clear rights soon after they receive music from labels.
Some DSPs are starting to require that record labels include information about at least one composition rightsholder for each track they submit. And many DSPs, particularly large ones with multi-territory footprints, have also negotiated direct licenses with major music publishers in which the parties work out arrangements for determining rights ownership and paying royalties.

But otherwise, it remains the case that DSPs often have neither matching composition nor complete split information for the music in their enormous and fast-growing catalogs.

IDENTIFIERS AND METADATA DON'T TRAVEL WITH ASSETS

One way to ensure that music is identified properly is to install identifiers and metadata into music files themselves when it is created. This is especially important in music files that end up "in the wild" or uploaded to services by end-users. Tools such as VEVA Collect and Downtown Music Group's Songspace enable artists to provide metadata for music assets as part of the creation process today.
But generally, files don't have identifiers or metadata installed, and even if they did in the first place, that information can disappear when files are copied or (especially) converted to other formats. One technique for installing identifiers into music files is digital watermark insertion,[12] but the industry hasn't adopted that to any significant degree.

Instead, the industry has found methods for identifying music files without accompanying metadata. The most widely used technology is acoustic fingerprinting, which is used in automated content recognition (ACR). A fingerprinting algorithm examines the bits of a music file and computes a series of numbers (the "fingerprint"), which it looks up in a database of known fingerprints to find a match. It is designed to compute the same fingerprint for different files that contain the same song, no matter the source or format. Although this technique is widely used today (e.g., YouTube's Content ID system uses it) and can be very accurate for identifying original sound recordings, these methods cannot differentiate between different versions of the same recording, which can be important for rights administration purposes. Fingerprinting is also sometimes used to detect musical compositions (e.g., cover versions); its accuracy for that purpose is significantly lower than for sound recordings.

The industry has found various ways to identify music without metadata, but they have drawbacks.

EXPLOSION OF USER-GENERATED MUSIC USES

As mentioned above, the uses of much on social platforms like TikTok are exploding. Yet while many musical artists have their own presences on these services, they are designed to give primacy to the users who use music to show something else (their dances, lip-synchs, etc.) rather than to the artists whose music they are using. And while libraries of music samples such as Songtradr and Tracklib exist that are licensed.

Every one of these new user-generated creations is potentially a royalty-bearing item for music rightsholders, yet the means of capturing information about all these uses and turning it into revenue are patchy at best. Some

Every piece of user-generated content with music is potentially a royalty-bearing item for rightsholders.

of these user-generated content platforms use fingerprinting methods to identify copyrighted music in uploaded clips. Other platforms rely on blanket licenses with major rightsholders. Still others take the position that it is not their responsibility but that of end users to license music properly for use on their platforms. Meanwhile, recent legislation in Europe (Article 17 of the 2019 EU Copyright Directive[13]) is putting more pressure on online services that accept user-uploaded content to take licenses to the copyrighted material in those uploads, but the practical impact of this legislation is not likely to be felt for years.

Making the most of the licensing opportunities inherent in these user-generated platforms requires metadata and licensing data that is accurately connected to the underlying music assets; and as the requirements for licensing mechanisms for these platforms becomes clearer, so will the need for metadata management and communication.

2 Economic consequences of data mismanagement

The economic consequences of poor metadata management can manifest themselves in a variety of ways. It is also clear that the digital music business will continue to require better metadata management as a driver of revenue growth. Here are a few generic examples; later we will see a few specific examples with their financial benefits.

Entities throughout the digital music supply chain can benefit financially from data collaboration.

Music publishers

A music publisher needs data covering all the royaltable aspects of compositions in its catalog. This involves maintaining a robust database of composition information including ISWCs, songwriters with their IPIs, splits, sub-publishers, CMOs, and so on, some of which is intended to be supplied in CWR files but may not be complete or accurate.[14]

It also means collecting and maintaining ISRCs and associated data for sound recordings of those compositions. Yet multiple recordings of the same composition will have different ISRCs, the publisher may not have information about all of them in its database, and more recordings that embody the composition (and derivatives of those recordings, such as clips used in social videos) can appear over time.

Catalog investment funds

An increasing number of entities, including private equity backed and publicly traded investment funds as well as traditional music publishers, are buying catalogs of compositions by legacy songwriters with a view to monetizing them more effectively than their current or previous owners.
If the owner of a catalog lacks good data about composition rights ownership and matching sound recordings, then it is difficult to value the catalog accurately before acquisition. It also diminishes opportunities to monetize the catalog once it has been acquired and causes delays before collection can begin.

An analogous situation applies to other acquisition scenarios, such as labels acquiring other labels.

Record labels

As discussed above, record labels and indie distributors have generally not managed information about songwriters or music publishers involved with compositions embodied in sound recordings in their catalogs. Yet there are several reasons why labels and distributors should be more interested in doing this. First, as mentioned above, some DSPs have started to require that labels' feeds of sound recording information contain at least some composition rightsholder data.

In addition, labels as well as publishers and catalog owners miss out on various licensing opportunities by not making accurate, current rights owner data available. For example, music supervisors in television, film, and game productions are more likely to take synch licenses to music that's easier (as well as cheaper) to license; this requires making good data available to synch licensing hubs as well as to individual licensees for easier search and discovery of licensable music. The ability to maximize synch revenue is of increasing importance to rightsholders now that synch is the fastest-growing category of revenue in the industry.

Finally, labels that focus on EDM have the special challenge of releasing music that contains lots of samples and interpolations,[15] and if those aren't cleared properly, the labels risk copyright infringement liability.

CMOs

A performing or mechanical rights organization often needs data that is more complete than what rightsholders feed to it in order to disburse royalties completely and accurately in response to usage data from DSPs, performance venues, radio stations, and so on.

Incomplete data can lead CMOs to rely on estimates to calculate royalty payouts, which reduces accuracy and often penalizes "long tail" creators.

It can also cause long delays in the processes that result in getting royalty payments from DSPs as the CMOs work to find rights ownership information.

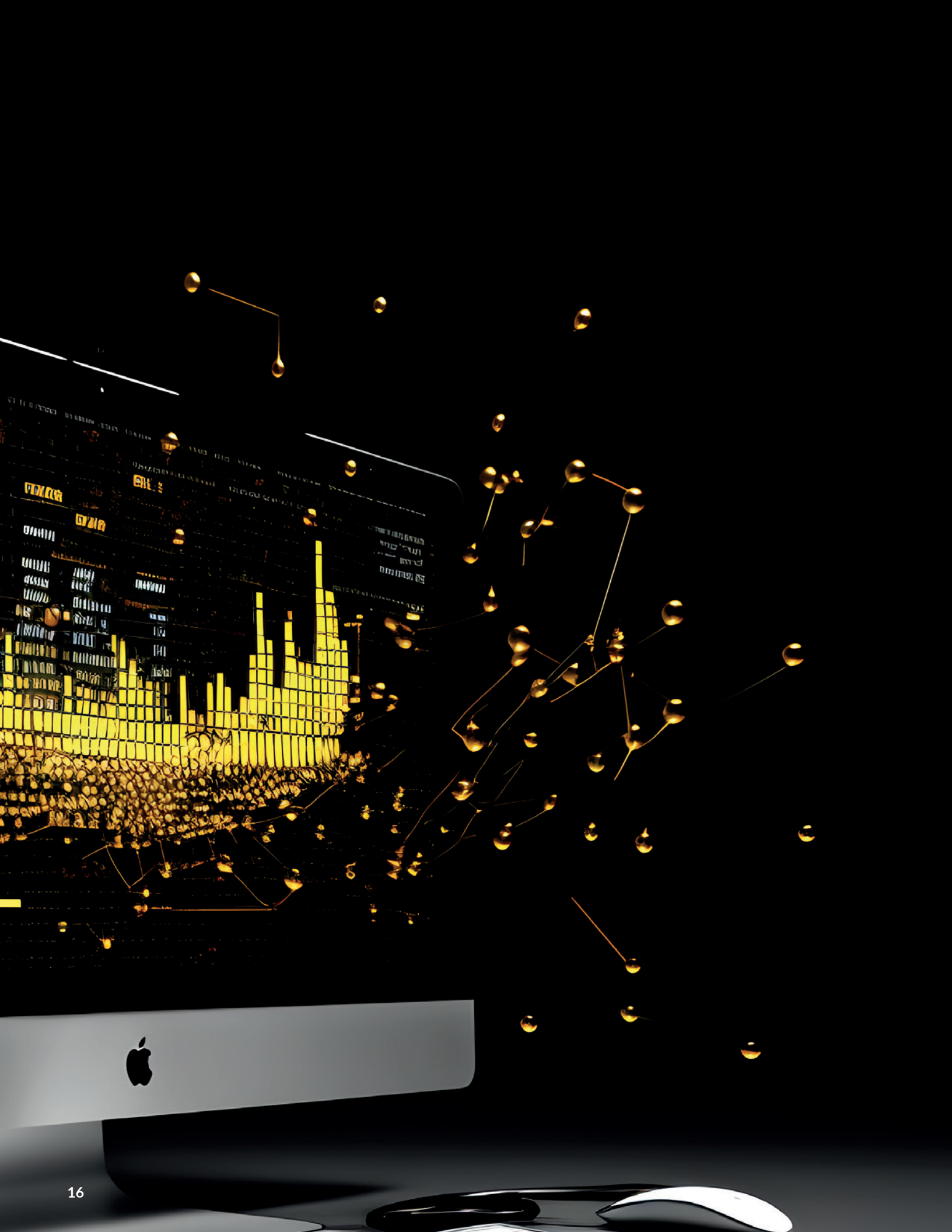
DSPs

DSPs face potential legal liabilities for playing music to which they haven't cleared the rights or for which they don't pay royalties properly. That's one reason why some DSPs outside the United States retain licensing agencies to do composition matching and rightsholder and split identification for them instead of relying on CMOs.[16]

DSPs also spend more time and effort in processing (and reprocessing) royalty claims and payments if they have data that is inaccurate, incomplete, or not delivered in a timely manner.

Finally, fraudulent activity on DSPs leads to fraudulent data that can have negative economic consequences. Some examples of fraud are well known, such as "bots" that inflate stream counts for artists, often using stolen DSP account credentials.

A study in early 2023 in France suggested that at least 1-3% of streams in that country, possibly much more, are fake,[17] while Deezer has stated that it identifies 7% of streams on its platform as fraudulent. Other types of fraud affect music supply chain data more directly, such as submitting unauthorized copies or remixes of tracks to DSPs.[18]



3 Data collaboration

The music industry is on a long-overdue path to better management of its data. Currently much of the industry uses a combination of proprietary data and messaging, some of which is based on standards such as those described previously in the paper. Most organizations either maintain their own proprietary data or rely on third party proprietary data sources.

The existing standards are excellent ways to automate and scale data transmission from one party to another. But they can sometimes be used in ways that are inconsistent across different entities. For example, these standards' syntaxes are used consistently, but their semantics can differ across implementation and the data carried in messages can contain errors and inconsistencies.

Starting around 2010, the industry explored the idea of building a global centralized database for music rights and rightsholder information that would be accessible through standard interfaces. This approach is attractive on a few levels: it can eliminate redundancies and inconsistencies, it can drastically lower rights administration costs for all stakeholders, and it can enable technology infrastructure and expertise to be deployed in cost-efficient and scalable ways. But when it was tried in a few different projects, this "Grand Unified Database" strategy ran into insurmountable problems with funding and governance before they could even tackle the unprecedented operational challenges inherent in such an enterprise, so none of them succeeded.[19]

The idea of a global centralized database of music rights information has been tried without success.

A more recent response to the need for good data has been for certain organizations to create databases that are openly accessible instead of closed and proprietary. For example, the MMA in the United States requires the Mechanical Licensing Collective to make its data available through a public website and to make full data feeds available to qualified parties; it does this today through bulk weekly downloads of its database containing information about more than 30 million compositions that it provides at a very low cost.
Even more recently, PRS, the UK's compositions CMO, announced a project called Nexus to provide an API accessible database of composition and matching sound recording metadata for DSPs to use.[20] In both cases, rightsholders can use the agencies' web portals to examine and submit corrections and updates to data about their own compositions.

The move towards open data goes beyond CMOs. Some DSPs, such as Spotify and Deezer, offer APIs (interfaces to their data for software developers and authorized companies) that enable querying of data from their music catalogs and related information such as playlists. Other entities make their databases available for searching through public websites, such as IFPI's ISRC database (powered by the US sound recording PRO SoundExchange) and the joint ASCAP-BMI database Songview. And open source databases such as Musicbrainz and Wikidata can often be useful as well.
Yet even if many more organizations make their data available openly, the overall effect will fall short of the type of data integrity and interoperability that would be possible with an idealized central database. Some organizations won't make their data available openly, and problems of data consistency and quality across organizations will remain.
That brings us to data collaboration. In data collaboration, multiple entities communicate data amongst themselves in order to improve its collective quality and usefulness to all entities for particular purposes; and there is no attempt to create a universal database. The entities engage services—either their own or from third parties—to enhance the data that travels among the entities in various ways.

Data collaboration is the next stage in music data management.

3 general types of enhancement through data collaboration are possible:

Matching:

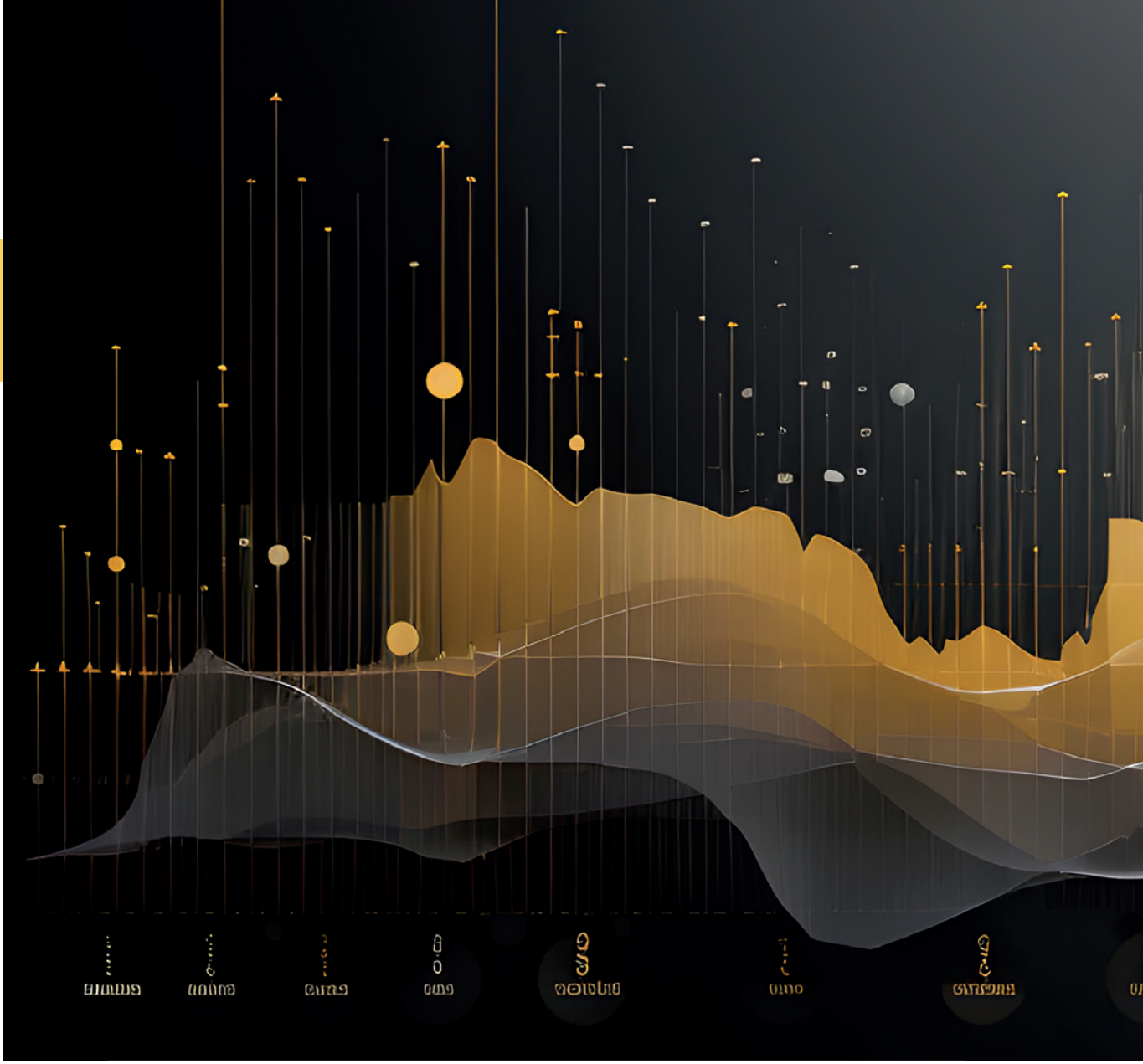
matching one entity's data to other types of data from the other entity in ways that improve each entity's ability to process transactions. For example, matching sound recordings from a record label to compositions in a publisher's catalog, and matching related information such as IPI numbers across entities.

Completion:

filling in one entity's missing data with data that another entity can provide. For example, completing rightsholder and split information for different CMOs that have only partial views of that information for compositions with rightsholders—such as when rightsholders span multiple territories—to help ensure accurate valuation of the publisher's catalog.

Maintenance:

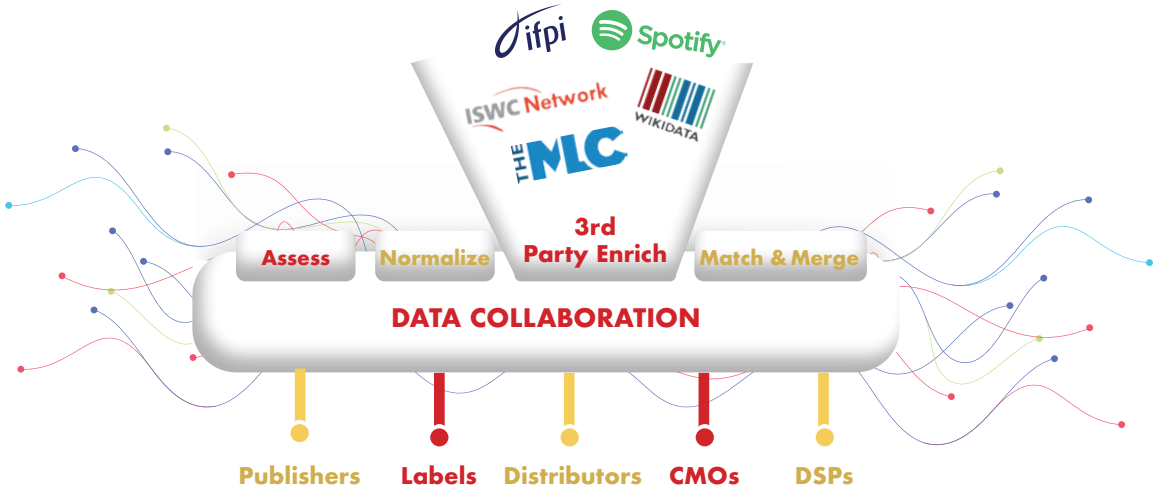
performing updates in an entity's own data as it changes over time to fix inconsistencies and record changes. For example, reflecting changes in splits or changing publisher affiliations to reflect ownership changes when catalogs are bought and sold; or adding sound recording information (such as ISRCs) to data about a composition as more recordings of it are created.



All of these types of enhancements also require that data be compared and assessed over time to ensure quality. It may not be necessary for every entity in the supply chain to have exactly the same data values, but the entities should have a methodology that gives them a way of assessing whether data consistency and quality is sufficient to be confident that it's usable for a given purpose.

By analogy, American banks rely on credit scores generated by the FICO methodology to judge would-be borrowers' creditworthiness, although each bank can determine the minimum credit score that qualifies a borrower for a certain loan. Maintenance also helps solve problems with inaccuracies that can result from errors in input, transmission, or format conversion, which are also major sources of improper royalty payments. And maintenance is the key to ensuring that data quality improves over time.

Data collaboration requires, at a minimum, that the entities involved each have ways of exchanging data with one another and with third parties, such as open databases, APIs, and standards.



The diagram in the figure above shows a general architecture of data collaboration in music. The types of operations involved in data collaboration can include:

Assessment: When multiple pieces of data across different entities should or might be identical but aren't, assessment determines whether they should be treated as the same data or not based on their degrees of similarity.

Normalization: Normalization subsumes loosely structured or unstructured data into a structured data scheme with standardized values.

Third-Party Enrichment: Using trusted third party data sources to supplement or validate existing data.

Matching & Merging: Finding data that completes or complements existing data and merging it with the existing data. For example, finding missing songwriters and their splits for a given composition, or finding additional sound recordings of a given composition.

Some of these steps require sophisticated algorithms, such as determining similarity and truth among large numbers of entities. Machine learning techniques can be used over time to develop rules, and/or human intervention may be necessary to make determinations where the automated rules fall short.

They also require careful consideration of permissioning of the data, so that each organization is able to share only enough data to make these operations possible and isn't compelled to give away their "crown jewels." In addition, the collaborative nature of these processes eliminates redundancies of each organization having to manage its own data completeness and accuracy operations—such as multiple CMOs and/or DSPs having to match composition and rightsholder data to recordings on very large catalogs that contain largely the same material.

The existing standards in the music industry certainly help enable these techniques, but they aren't sufficient by themselves;

for example, there is no widely adopted standard for announcing changes to data other than for recording releases via the DDEX ERN protocol (where "new" data about a sound recording supersedes the "old" data).

Data collaboration technologies have been deployed successfully in many industries outside of music, using technologies from vendors such as Snowflake, Teradata, and Amazon Redshift.

Blockchain technologies have also been used to enable data collaboration and consensus in cases where the scale is appropriate. The significant business benefits of data collaboration have been quantified in surveys.[21]

4 Financial benefits of data collaboration

Enhancements from data collaboration lead to real financial benefits for all supply chain participants.

For each of the three above types of data enhancement, here are examples drawn from real-world situations showing models for calculating potential quantitative benefits that are hypothetical yet realistic:

Matching: recordings data for music publisher

Pioneer Music Publishing (PMP), a major music publisher, has a catalog of 100,000 compositions that generate an average of $300 per year from digital revenue, for a total of $30 million per year.[22]

For each composition (ISWC), PMP collects revenue from only some of the sound recordings (ISRCs) that exist. Assume that PMP only collects on one-third (33%) of the ISRCs that exist for each ISWC; and the percentage is shrinking as the number of sound recordings embodying each composition grows with the explosion of user-generated content, such as on mashups, remixes, and short-form videos.

Assume further that 20% of the compositions in PMP's catalog generate 80% of the revenue, i.e., that these 20,000 compositions generate $24 million or $1200 per composition.

PMP shares data with record labels and other sources, and gets information that enables it to collect on half the remaining ISRCs for the 20,000 high-value compositions, i.e., the number of recordings on which it can collect doubles.

Assume that the newly discovered ISRCs come mainly from social media uses of content which generate only $300 (instead of $1200) annually per composition. Still, this increases PMP's digital revenues by $300 times 20,000 or $6 million, which is an increase of 20%.

PMP can repeat this process on a regular basis to capture even more incremental revenue from new user-generated uses of their content.

A music publisher can increase revenue through data collaboration by finding more recordings of compositions in its catalog.

COMPLETION: CATALOG DATA ENHANCEMENT FOR ACQUISITION

Hellrazor, a British metal band that had a string of classic albums in the 1980s, is putting its catalog of compositions up for sale. Hellrazor's management wants to show the maximum possible value of the catalog to potential buyers, which include private equity-backed catalog investment funds.

They have information about the band's own recordings and some others (e.g., tribute bands and samples used in hip-hop tracks) but not about uses of their music in TikTok videos or elsewhere.

Hellrazor's catalog includes only 70 compositions from 7 studio albums, but they pull in an average of $10,000 in digital revenue annually, once again with 20% of the catalog generating 80% of the revenue.

A composition catalog investment fund can increase revenue through data collaboration by finding uses of its compositions in user-generated videos.

They have information on an average of 4 recordings (ISRCs) per composition (ISWC). Through data collaboration, Hellrazor's management is able to find an average of one more ISRC per high-value composition (25% increase in ISRCs), from new sources such as TikTok videos and more samples, and include that data in the information it makes available to potential buyers.

This enables the band to increase its potential valuation by as much as 20% (80% of 25%), depending on the revenue generated by those new ISRCs.

It also enables the buyer of Hellrazor's catalog to start collecting on a much larger number of recordings associated with the compositions, and to claim and collect more quickly than if it had to research the information itself after the acquisition.

MAINTENANCE: COMPOSITION DATA ENHANCEMENT FOR REGIONAL DSP

Swara is a streaming music service based in Jakarta that serves Southeast Asia and has a 20% overall market share in countries with a total population of almost 700 million.

Swara has global deals with the major labels and music publishers, and it gets feeds from many indie distributors. But in some of the countries where Swara operates, it is unable to rely on the local CMO to clear rights and claim royalties for compositions, so the processes for paying composition royalties for repertoire outside of the major labels and publishers are unclear or nonexistent.

Indie publishers can share data with DSPs to increase royalties in certain territories.

CIMP (Coalition of Independent Music Publishers) is a trade association of indie music publishers based in London.

CIMP is working with its member publishers to increase royalty payments in territories outside the UK, North America, and Europe by sharing composition rights data with DSPs in those territories.

CIMP organizes data collaboration among its members, songwriters, and third-party sources to provide more comprehensive rights ownership data for compositions and matching ISRCs to large DSPs such as Swara on a continuing basis over time, so that the DSPs always have the latest and highest quality data.

Swara participates in the data collaboration so that it can pay composition royalties with confidence; it can lower its risk of legal liability for uncleared composition rights, avoid having to take songs down or refuse submissions from certain independent labels, and reduce the overhead time and cost of clearing rights that the local CMOs are unable to clear.
Meanwhile, CIMPs members could benefit from as much as 25% increased revenue from the territories where Swara operates.[23]
These examples all show revenue increases for rightsholders that are reasonably expected to be in the 20-25% range.

Therefore it's fair to say that the average financial benefit for rightsholders under these scenarios is around 20% annual revenue. Data collaboration also benefits other stakeholders, such as CMOs and DSPs, through decreased legal liability risk, increased efficiencies in royalty processing, and lowered operational costs.

5 Conclusion

Data collaboration is the key to increasing revenue for music industry rightsholders through the explosion of user-generated content as well as through more revenue collections from existing streaming activities and new licensing opportunities in areas such as synch.

Data collaboration involves exchanging data with other entities in the supply chain, possibly with the help of service providers that can clean, enhance, and coordinate data continuously over time. It builds on the industry's existing foundation of standard identifiers, messaging, and metadata schemes, and it takes advantage of the growing trend towards APIs and open databases, while avoiding the pitfalls of the single centralized database approach.

As we've shown here, data collaboration addresses many of the economic consequences of structural data problems in the music industry, such as the lack of complete and accurate rights data associated with assets throughout the digital supply chain—problems that lead to more money being left on the table as the supply of music, and the diversity of user-driven uses of music, continue to increase globally. We have shown instances of how data collaboration can lead to significant quantitative as well as qualitative benefits, with financial benefits that can be expected in the 20% range for many stakeholders.

Glossary

API	Application programming interface. A way of accessing a system through routine calls in software.
BIEM	Bureau International de l'Edition Mécanique, the international organization for MROs.
CCID	Claim Confirmation and Invoice Details, a CISAC file format for CMOs to send information to DSPs about claimed royalties.
CDM	Claim Detail Message, a DDEX standard for CMOs to send information to DSPs about claimed royalties.
CISAC	Confédération Internationale des Sociétés d'Auteurs et Compositeurs, the international association for authors' and composers' CMOs.
CMO	Collective management organization. A collective rights management entity that can license performance and/or mechanical rights. PROs and MROs are types of CMOs.
CRD	Common Royalty Distribution, a CISAC file format used for communication of composition usage among CMOs and music publishers.
CWR	Common Works Registration, a CISAC file format used by music publishers to send information about compositions to CMOs.
DDEX	Digital Data EXchange, a family of music messaging protocol standards, also the name of the standards body that manages them.
DSP	Digital service provider. A consumer-facing digital music service.
DSR	Digital Sales Report, a DDEX standard for reporting sales or usage of music tracks to rightsholders or administrators.
EAN	European Article Number, also known as International Article Number, a standard identifier for physical products closely related to UPC.
ERN	Electronic Release Notification, a DDEX standard for feeding record release information to DSPs.
IFPI	International Federation of the Phonographic Industry, the international umbrella trade association for the recorded music industry.

IPI	Interested Parties Information, a unique identifier for songwriters and music publishers.
ISNI	International Standard Name Identifier, a unique identifier for names, often used for recording artists.
ISRC	International Standard Recording Code, a unique identifier for sound recordings.
ISWC	International [musical] Works Code, a unique identifier for musical compositions.
MLC	Mechanical Licensing Collective, the U.S. MRO established as part of the MMA.
MMA	The Music Modernization Act of 2018, United States legislation that established the MLC.
MRO	Mechanical rights organization. A type of CMO that licenses mechanical rights to musical compositions.
PRO	Performing rights organization. A type of CMO that licenses public performance rights to music. PROs for sound recordings are sometimes called neighboring rights organizations.
UPC	Universal Product Code, a standard identifier for physical products closely related to EAN.

About verifi media

Verifi Media is a global leader in modern media rights data management services focused on improving comprehensive music rights data, thereby empowering media creators, owners and representatives through digital data innovation.

Using modern tools like cloud computing, machine learning techniques and algorithms, Verifi revolutionizes how media ownership and meta data is enhanced, corrected, shared and tracked across the supply chain, resulting in significantly better business decisions while enabling creators to be paid properly for their work.

Verifi Media currently works with many of the largest global music companies in the world, including members of the Verifi Rights Data Alliance.

For more information, visit www.verifi.media.

About the Author

Bill Rosenblatt is the founder and president of GiantSteps Media Technology Strategies, a consulting firm that has provided expertise in technologies, business strategy, and intellectual property for digital media since 2000.

His music industry clients have included major record labels, digital music services, technology companies, investment firms, and startups around the world. He has advised public policy entities in the United States, Europe, and elsewhere on digital copyright issues. He has served as an expert witness in litigations and royalty rate proceedings on the digital music market, digital copyright, and related technologies.

Bill is an adjunct professor in the Music Business program at New York University. He is the program chair of the annual Copyright + Technology conferences in New York, which he co-produces with the Copyright Society, and he has spoken at other conferences worldwide including the World Economic Forum in Davos, Switzerland. He is on the trustee boards of the Copyright Society and Princeton Broadcasting Service, Inc.; and he is on the advisory board of the American Society for Collective Rights Licensing, a collecting society for illustrators and photographers.

Bill is the author of Digital Rights Management: Business and Technology (Wiley, 2002), coauthor (with Howie Singer) of Key Changes: The Ten Times Technology Transformed the Music Industry (Oxford University Press, 2023), and a media industry contributor to Forbes.com. He holds degrees in computer science from Princeton and the University of Massachusetts.

FOOTNOTES

[1] IFPI Global Music Report 2023 – State of the Industry. Available at https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/GMR_2023_State_of_the_Industry_ee2ea600e2.pdf.

[2] IFPI Global Music Report 2023.

[3] Goldman Sachs, Music streaming services are on the cusp of major structural change, https://www.goldmansachs.com/intelligence/page/music-streaming-services-are-on-the-cusp-of-major-structural-change.html.

[4] https://luminatedata.com/reports/midyear-music-industry-report/

[5] IFPI 2022

[6] https://newsroom.tiktok.com/en-us/year-on-tiktok-music-report-2021.

[7] https://www.cisac.org/.

[8] Murray Stassen, There Are Now 120,000 New Tracks Hitting Music Streaming Services Each Day, Music Business Worldwide, May 25, 2023, https://www.musicbusinessworldwide.com/there-are-now-120000-new-tracks-hitting-music-streaming-services-each-day/.

[9] The early streaming services owned by major labels, such as pressplay and MusicNet, cleared mechanicals in a deal with the National Music Publishers Association, but independent streaming services had no such arrangements.

[10] For example, overlapping claim files from multiple sources could claim more than 100% of rights in a given composition (known as "overclaiming").

[11] Variations can include quirks in a country's copyright laws, multiple CMOs in the country, CMOs that operate in multiple territories, DSPs paying so-called special purpose vehicles (SPVs) that represent Anglo-American repertoire, and so on. Another variation is performance royalties in the U.S., which DSPs pay to PROs on the basis of blanket licenses based on shares of catalog rather than via a claiming system.

[12] Watermarks that are meant to be inaudible can only hold enough data to store an identifier, but that identifier can serve as an index to a database that stores metadata.

[13] Article 17 of the EU Copyright Directive https://eur-lex.europa.eu/eli/dir/2019/790/oj.

[14] For example, a composition may not have an ISWC assigned if songwriter splits are not defined when a sound recording embodying that composition is released, or if different rightsholders register the composition with conflicting information about identities or splits.

[15] An interpolation is a re-creation of a melody and other elements of a musical composition, as opposed to a sample, which is a copy of a segment of a sound recording. An interpolation usually requires only a license for the composition, not a sound recording embodying that composition. A well-known example of interpolation is Afrika Bambaataa's "Planet Rock," which contains an interpolation of Kraftwerk's "Trans-Europe Express."

[16] In the United States, the Music Modernization Act of 2018 removed legal liability for improper payment of mechanicals for participating DSPs.

[17] Centre national de la musique, Fake streams, real phenomenon: the CNM working with the industry to fight streaming fraud, January 16, 2023, https://cnm.fr/wp-content/uploads/2023/01/CP_CNM_Manipulation-des-streams_ENG.pdf. Spotify, Deezer, and Qobuz provided data for the study.

[18] See Elias Leight, Why Can't Music Fix Its Fake Streams Problem?, Billboard, April 5, 2023, https://www.billboard.com/pro/fake-music-streams-industry-fraud-problem/.

[19] The best known of such attempts was the Global Repertory Database (GRD), which was first announced in 2010 and was abandoned four years later after more than $13 million was spent.

[20] Presentation at https://www.wipo.int/meetings/en/details.jsp?meeting_id=76608.

[21] See for example: Laurence Goasduff, Data Sharing Is a Business Necessity to Accelerate Digital Business, Gartner, May 20, 2021, https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business.

[22] This and subsequent numbers are estimates based on publicly available financial data from large rightsholders.

[23] Assuming that Swara was not paying composition royalties on any of CIMP members' compositions; because it has 20% market share, Swara represents 25% more subscribers than the other DSPs serve in Southeast Asia.