

## Article

# A Proposed Settlement and Distribution Structure for Music Royalties in Korea and Their Artificial Intelligence-Based Applications

Youngmin Kim <sup>1,\*</sup>, Donghwan Kim <sup>2</sup>, Sunho Park <sup>3</sup>, Yonghwa Kim <sup>1</sup>, Jisoo Hong <sup>1</sup>, Sunghye Hong <sup>1</sup>, Jinsoo Jeong <sup>1</sup>, Byounghyo Lee <sup>1</sup> and Hyeonchan Oh <sup>1</sup>

<sup>1</sup> Korea Electronics Technology Institute, Seoul 03924, Republic of Korea; yonghwa.kim@keti.re.kr (Y.K.); jhong@keti.re.kr (J.H.); shhong@keti.re.kr (S.H.); jjeong@keti.re.kr (J.J.); yui145263@keti.re.kr (B.L.); hyeonchan.oh@keti.re.kr (H.O.)

<sup>2</sup> Catbell Company, Seoul 06526, Republic of Korea; stan@catbellcompany.com

<sup>3</sup> Chorock Soft, Seoul 06123, Republic of Korea; doctorf@chorock.co.kr

\* Correspondence: rainmaker@keti.re.kr; Tel.: +82-2-6388-6695

**Abstract:** Digital music is one of the most important commodities on the market due to music royalty distribution in Korea. As the music market has been transformed into a digital music market by means such as downloading and streaming, the distribution of music royalties via online service providers (OSPs) has become a highly important issue for music rights holders. Currently, one of the most important issues in music royalty distribution in Korea is the unfair distribution of royalties due to the indiscriminate repeat streaming of digital music. To prevent this, music consumption log data from several OSPs were collected via a day-based system; however, there was a limit on the identification of detailed information on the use of music in its current state. This paper analyzes the structural problems and limitations related to the settlement of music royalties and provides a structure in which there can be transparent settlement and distribution between users and rights holders as an institutional measure. We also propose various AI (artificial intelligence)-based applications using music consumption log data. The proposed system will hopefully be used for public purposes.

**Keywords:** music sound source; copyright information; log data; artificial intelligence; database; settlement; distribution



**Citation:** Kim, Y.; Kim, D.; Park, S.; Kim, Y.; Hong, J.; Hong, S.; Jeong, J.; Lee, B.; Oh, H. A Proposed Settlement and Distribution Structure for Music Royalties in Korea and Their Artificial Intelligence-Based Applications. *Appl. Sci.* **2023**, *13*, 11109. <https://doi.org/10.3390/app131911109>

Academic Editor: Andrea Prati

Received: 31 July 2023

Revised: 26 September 2023

Accepted: 26 September 2023

Published: 9 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Due to the development of streaming-based web services and mobile devices, the modern music industry has become digital. Unlike in the past, when music chart rankings were based on the offline sales of cassette tapes and CDs (compact discs), the numbers of downloads, streams, and media appearances are now becoming more important factors [1–4]. Compared to the analog method based on offline sales, the digital method was once expected to provide a clear and efficient alternative to music royalties, but in fact, it is difficult to claim that the settlement or distribution is transparent and efficient in Korea for a few reasons.

First, as the system transitioned from analog to digital, the music industry also made attempts to digitize existing music (UCI issuance). However, each online service provider (OSP), representing one of the dominant elements in the Korean music industry, issued and used a different UCI, so it was not easy to unify UCIs.

Second, there are no institutional procedures for transparent settlements. For the transparent settlement of music, it is necessary to accurately collect data regarding how much music has been downloaded and streamed. Such data have been collected accurately for individual OSPs because they provide music services; however, OSPs are reluctant to disclose information such as the number of subscribers and their related information. After

several incidents and lawsuits related to music royalties, the Ministry of Culture, Sports, and Tourism, the Korea Copyright Commission, and the Korea Music Content Association discussed guidelines for a log collection system in 2013 [5]. As a result of this discussion, each OSP must provide log information to the Korea Copyright Commission every day. However, the log collection system needs to be improved because of unclear management.

Third, each association can reap unsettled dividends because of unclear sales information. In completely securing log collection data, OSPs incur additional costs, which may cause unnecessary disputes. In fact, the transparent disclosure of data increases the workload of data generators and collectors, such as trust service providers or OSPs, and there is a lack of manpower for managing errors in data quality.

In this study, we analyze this issue and suggest the following solutions for solving these problems and protecting music users, rights holders, and public interests: (1) analyzing the structural problems and limitations of calculating and settling music usage; (2) providing a structure for transparent settlement and distribution between users and rights holders; (3) assuming several methods based on artificial intelligence technology to utilize music log data for public purposes; and (4) providing alternatives to regain transparent copyright and trust in the music industry by presenting plans for future legislation.

## 2. Collection of Music Usage Information and Analysis of Structural Limitations

There has been considerable progress regarding the distribution and settlement of music royalties abroad. DDEX (Digital Data Exchange, LLC, New York, NY, USA), which is an international standard focused on the creation of digital value chain standards to make the exchange of data and information across the music industry more efficient, is one of the best examples [6]. It has been recognized that the DDEX standard is a protocol among interested parties such as copyright holders, music publishers, distributors, metadata companies, music licensing companies, studios, and archives. The DDEX value chain consists of several standard regulations. First, music release notifications and licensing regulations are classified via musical work notification (MWN), musical work licensing (MWL), and letters of direction (LoD). Second, the DDEX standard that communicates information about sales, usage, or revenue generated on a digital music retailer service to the relevant rights owners is the Digital Sales Report Message Suite Standard (DSR), which is known as the Sales and Usage Reporting standard. The DSR standard, which is in a flat-file format, is used to enable the reporting of information regarding the level of sales, usage, and/or revenue generated from the distribution of music or videos, along with information about that music, in order to enable rights owners to pass on royalty payments as appropriate. Third, the Claim Detail Message Suite Standard (CDM) provides three formats that support the owners/administrators of musical works and senders of sales and usage reports (usually DSPs) in the exchange of information about claims the owners/administrators of musical works make against DSPs with respect to the musical works that appear in the sales and usage reports. Fourth, the Electronic Release Notification Message Suite (ERN) enables record companies or distributors to send data to digital service providers (DSPs) that describe the releases and resources, such as sound recordings and music videos the record company or distributor is making available to the DSP for use on its service, and the terms and conditions under which such releases and resources can be made available to consumers. The most recent version of the ERN to be published is ERN 4.3. Fifth, the metadata rights and recording standard is a regulation between a music licensing company and a recording company. Finally, the metadata provided in a recording information notification (RIN) message are an integral part of the operational infrastructure across the music value chain and facilitate the process through which all rights holders and other contributors are properly remunerated. Such metadata can also enable retailers to provide better metadata about their products to consumers, which helps to stimulate search and discovery. Since the DDEX standard was launched in 2006, most music companies and distributors have followed this protocol; however, the situation is different in Korea for several reasons.

In Korea, copyright trust service providers have authority over settlement and distribution since they manage the rights delegated by copyright holders, music record producers, or performers. For example, when they carry out the settlement and distribution of music revenues to copyright holders, the copyright trust service provider sums up the music sales volume for a certain period of time (e.g., three months) and provides the data in the form of a report to each copyright holder. In this case, it is difficult to determine the settlement details because the settlement reports do not include accurate information, and there are a number of complex products from OSPs in the reports. In other words, the OSPs calculate the music sales for each user ID (not each music song) and then settle and distribute the music sales and profits to the copyright holders (in this case, copyright trust service providers). So, there are evitable errors between the music sales profits derived from a user ID and the royalties from the song(s). This asymmetry and discontinuity of information make transparent distribution and settlement difficult.

To solve this issue, the Ministry of Culture, Sports and Tourism, Korea Copyright Commission, and the Korea Music Content Association discussed guidelines for log collection, and in 2013, it was decided that each OSP must provide log information to the Korea Copyright Commission daily. However, the log collection system was based on an MOU (Memorandum of Understanding); there is no effective way to obtain log information or prohibit the suspension of transmission. It was agreed that the log collection system would provide information to the music ranking chart website for statistical use, and the settlement and distribution of music royalties would be performed using separate systems of associations. Table 1 is an example of a log item standard field.

**Table 1.** An example of a log item standard field.

Item	Example	Remarks
UCI code	uci:i500-12345678-1234567890-1	-
Granted code	S1234	From UCI
Sales year/month/day/time	20121001-13:05:30	Sales YYMMDD-HH:MM:SS
Service site	<a href="http://www.melon.com">http://www.melon.com</a> (accessed on 31 July 2023)	-
Song ID	0258576	Self-categorization
Song	Back in time	Self-categorization
Album code	2031	-
Album name	The moon embracing the sun	-
Artist name	Lyn	-
Service code	D-50	Standardization code from UCI
Distributor name	Recording Industry Association of Korea	-
Distributor song code	A101629692	-
Release day	20120215	Release YYYYMMDD
Producer name	Pan Entertainment	Agency or producer name
ICN code	ICN.201 = 1000077661-A101629692	-
ISRC code	KR-A01-99-00137	-

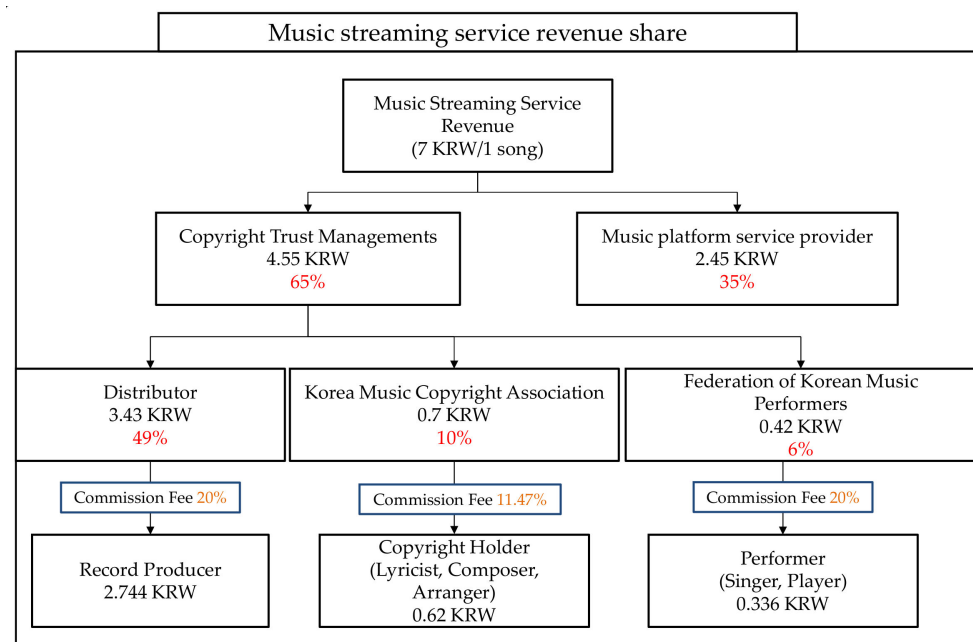
As shown in Table 1, there are five essential elements (UCI code, song ID (from OSP), service code (download/streaming), sell time, and ASP name) for the settlement and distribution of music royalties. According to the regulations on the collection of music royalties of the Korea Music Copyright Association, transmission royalties are defined as the higher output of the following usage formulas using the relevant information [5]:

$$\begin{aligned}
 \text{Royalties} &= (0.7 \text{ KRW (South Korean won) (rate per song)}) \times (\text{number of uses}) \times (\text{share or 700 KRW per month} \\
 &\quad (\text{rate per subscriber})) \times (\text{number of subscribers}) \times (\text{management ratio}) \\
 &\quad \text{or} \\
 \text{Royalties} &= (\text{Sales}) \times (10.5\% \text{ (music royalty rate)}) \times \text{management ratio} \\
 &\quad (\text{For the Korea Music Copyright Association, the management ratio is 96\%})
 \end{aligned} \tag{1}$$

Considering this, it seems as if the system could be operating plausibly for settlement, since music usage log data are collected based on an MOU and transmission royalties are derived according to Formula (1). However, the above system for the accurate settlement and distribution of royalties has a few deficiencies.

First of all, there is an inherent problem with the log collection system. The music usage log information, which is transmitted from major OSPs in Korea on a daily basis based on the MOU without institutional support by law, was originally obtained for statistical use. Therefore, regarding the issue of settlement, major OSPs deliver the transmission royalties to each association based on the above formula and announce sales data. Each association distributes the royalties based on the number of music rights holders. However, the transmitted music usage log data from OSPs are song-based sales data; the information collected from the log collection system shows how many times certain music is streamed or downloaded in a certain period, not indicating which music the users listened to or how many times they did. Therefore, it is highly difficult to confirm whether the data are accurate, since actual sales data are collected by monthly payment per user ID. Moreover, only major domestic OSPs participate in the MOU, while global operators such as Google, Apple, YouTube, and Spotify (accessed on 31 July 2023) do not participate. However, as they continue to grow as major OSPs, the music royalties are distributed appropriately by each global operator according to their own log data.

The second problem originated from the profit distribution structure that is focused on large production companies or OSPs. Figure 1 shows an example of the profit distribution structure of streaming music. When the streaming music royalty per song is KRW 7, the royalty of the copyright holders (lyricist, composer, and arranger) will be KRW 0.62, while the OSP and the Korea Music Copyright Association will receive KRW 2.45 and 0.7, respectively [7]. In other words, a substantial amount of the revenue from music sales is likely to be given to major OSPs or famous music record producers (intermediate distributors). Therefore, individual music copyright holders rely on the copyright trust service providers, meaning that the system of transmission royalties can be manipulated [8].



**Figure 1.** An example of music streaming service revenue share. (Red: revenue share)

The third problem may arise from the Personal Information Protection Act [9]. When an individual rights holder requests a detailed report regarding the transmission of music royalties to the association, which has authority over the settlement and distribution of royalties from individual rights holders, the report is requested by the association from

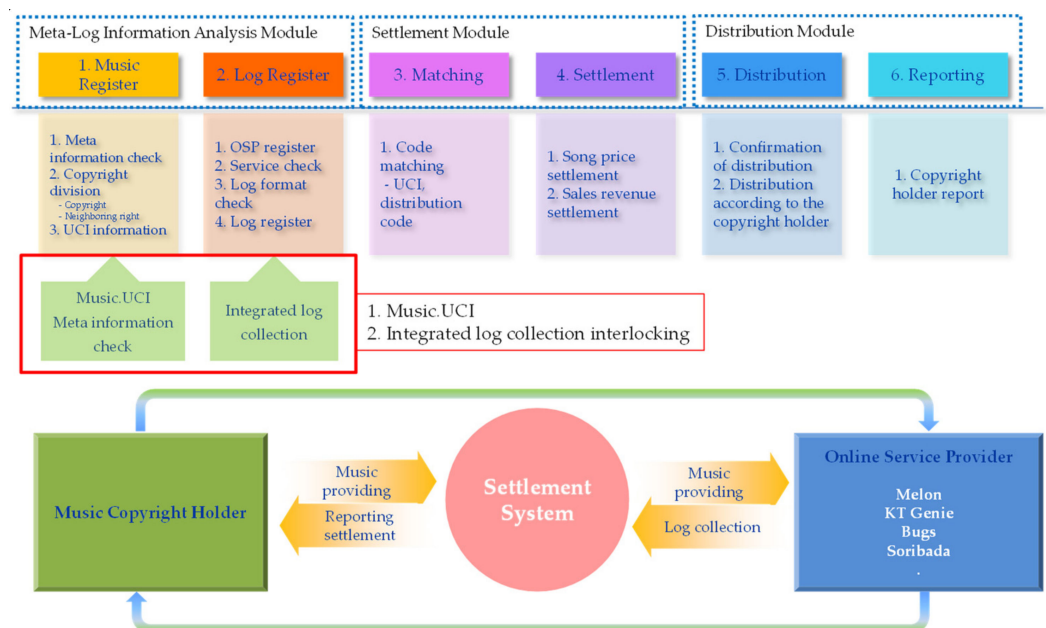
major OSPs. However, there are a number of complex music product sales in OSPs, and so an individual rights holder cannot expect an accurate, detailed report. The music sales information, number of subscribers, and subscriber details are key elements that are related to major OSPs' profits; thus, they may not provide such data accurately. Currently, music usage information is not disclosed to the general public and researchers, nor is it disclosed in detail to copyright holders [5].

The aforementioned problems could be solved by forcing an amendment to the law, ensuring that sales are calculated by obtaining both the subscriber and PG information of OSPs. However, this information is an OSP's key business information that is related to their profits, and so many different interests are entangled within it.

### 3. Suggesting a Settlement–Distribution Structure for Music Usage

This study suggests the following alternative structure using the existing log collection system to settle and distribute royalties for music usage. Because music rights holders have the right to access the music usage status and information of rights holders, this structure could be one of the efficient ways to carry out the transparent settlement and distribution of royalties. This structure is basically similar to DDEX [6,10]; however, it is not a legally binding force.

Figure 2 shows the proposed settlement–distribution structure based on meta-log information analysis. This structure consists of three modules: a meta-log information analysis module, a settlement module, and a distribution module. These modules could be successfully constructed based on spontaneous requests from music rights holders.



**Figure 2.** Proposed settlement–distribution structure based on meta-log information analysis.

First, the meta-log information analysis module performs the registration of the copyrighted music and log information. Only the music copyright holder, who has authorized music usage status and information, can register the copyrighted music itself for settlement and distribution. Once the music is registered, the basic metadata can be obtained from the music UCI data of the Recording Industry Association of Korea. The rights for a single music source can be divided into copyright and neighboring rights (production rights and performance rights) [5]. The current log collection system gathers the sales month, OSP classification, the song code and service code according to the OSP, the association type, UCI, and sales volume data. For a settlement, the rate per song and the reported settled sum for a specific OSP (the rate per song multiplied by the number of times that the music is used) will be needed. This is a highly complicated issue because the process

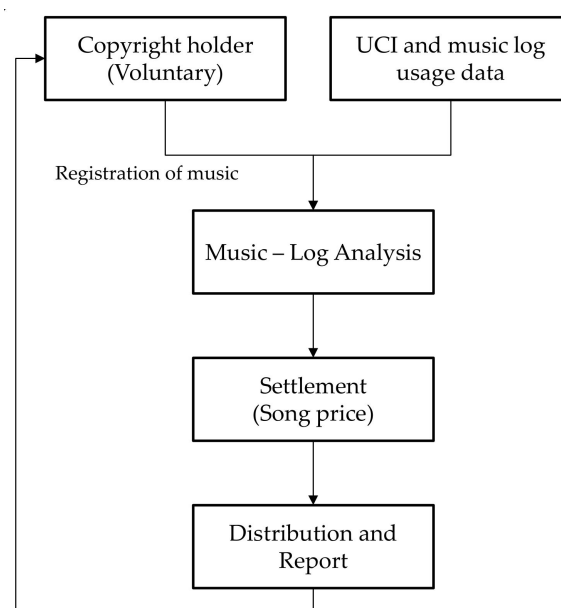


closely involves stakeholders (e.g., OSPs); however, the music copyright holder can request this information for their rights. Once the issues among the aforementioned stakeholders are solved, the proposed structure can be used to automatically retrieve the music usage log data from the log collection system, as long as the music rights holders register their copyrighted music.

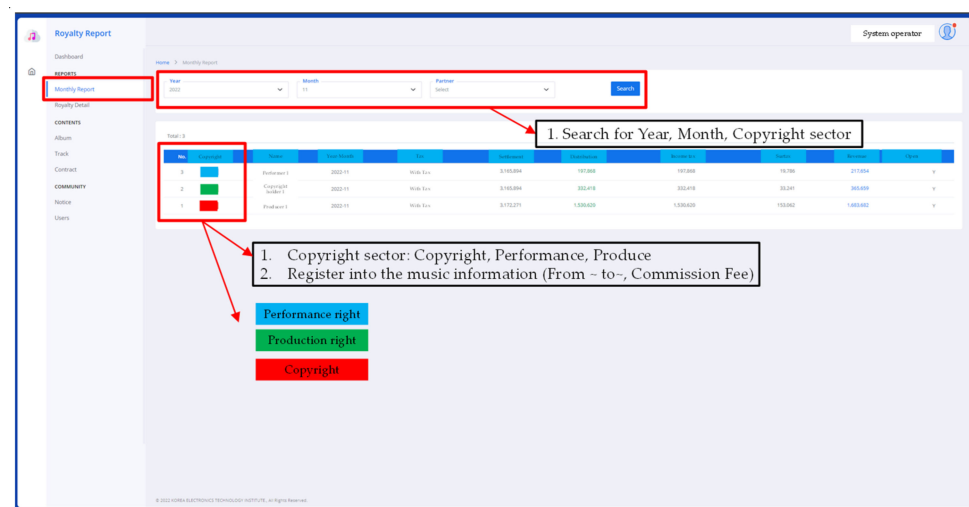
The second step consists of a match module and a settlement module. The match module carries out a matching process between the copyrighted music via registration and the music usage log data from the log collection system. The settlement process can be carried out based on the rate of a song or the share of total sales of a song, according to Formula (1). The share of total sales of a song is higher than that of the rate of the song in most cases, and this information can be requested by copyright holders because of their rights, while the information from major OSPs is confidential. Recently, several OSPs (e.g., NAVER VIVE (Seoul, Republic of Korea), etc.) adopted the settlement method of paying for music usage so that the fee would go directly to the music rights holders.

The third step consists of a distribution of royalties and a report system. When the settlement for distribution is finalized, the distribution ratio will be determined according to the copyright and neighboring rights in the first step. Based on this ratio, the royalties for the music's usage will be distributed to individual rights holders. The reporting system can provide a sum of royalties, royalties from different OSPs, and other information in a chart or a diagram.

Once the music rights holder registers the copyrighted music and each OSP transmits the log information for settlement, one can realize a transparent settlement and distribution structure because the music rights holders have the right to request sales details. Figure 3 shows a flow chart of settlement and distribution. After following this process, a settlement and distribution report can be drawn. Figure 4 shows an example of a settlement and distribution system constructed by applying these three modules. The royalty report allows the copyright holder to view the year, month, and copyright classification (copyright, performance rights, and production rights). We assumed that the metadata that are used in the system are generated via random generation from sample log data, and the song price settlement method was applied because music usage log data are not disclosed to the public.



**Figure 3.** Flow chart of the settlement and distribution of music royalties based on the proposed method.



**Figure 4.** An example of a settlement and distribution system applying all three modules.

#### 4. Applications Utilizing Music Usage Data Based on AI

In this section, we will explain the pseudonymization techniques of music usage information for protecting personal information in music usage data, and the applications utilizing music usage data based on AI will be introduced.

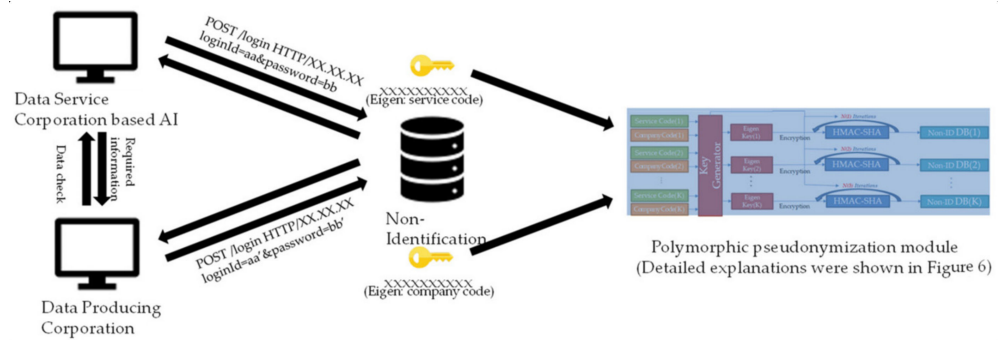
##### 4.1. Pseudonymization of Music Usage Information

Music usage information can be used for various purposes other than settlement or distribution. The usage information data in various fields have considerable inherent asset value, as in the case of MyData services in the finance field. However, in the music industry, the utilization of music usage information is limited due to the Personal Information Protection Act. Music rights holders give authority over their music rights to associations, and they permit the use of music-by-music distributors. Therefore, the personal information for the music is fragmented, and no one can carry out applications using the music usage information. However, in the proposed system, an agreement, such as usage consent, is previously constructed by registering the music in the first step, so that it will be immune from the legal issues. For the effective protection of personal information, it is more convenient that the pseudonymization of the music usage data be performed. This chapter discusses the ways in which public services or OSPs could utilize music usage data through pseudonymization before AI-based applications.

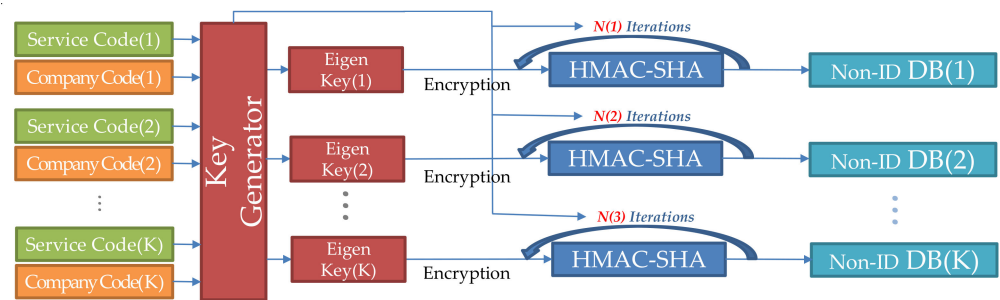
Music usage data contain personal information that is necessary to de-identify. The de-identification of personal information refers to the deletion or modification of part or all of this personal information so that a specific individual cannot be identified. The related guidelines in Korea provide five methods: pseudonymization, aggregation, data reduction, data suppression, and data masking. These five methods can be divided into anonymization and pseudonymization, depending on the deletion of personal identification. Anonymized information can be freely utilized because of the impossibility of re-identification; however, it lacks utility. On the other hand, pseudonymization data do not allow one to recognize individuals without additional information. Therefore, it can be used for purposes such as statistics, research, and the preservation of records of public interest [11,12].

The pseudonymization of music usage information could be undertaken by the OSP that produces the data; however, the popular method is to choose an independent organization that can deal with pseudonymization through the de-identification of data. Figure 5 shows a process diagram for managing the unique key of an independent organization. The transmitted music usage data are pseudonymized using a JAVA-based, high-speed, one-way pseudonymization module. Here, the pseudonymization algorithm needs two key functions. First, since the amount of day-based music usage data is large, it is easy to process pseudonymization only when the computing time of the algorithm is

extremely short. Second, it is necessary that the same identifier be used (for continuity) for the previous and current data. This is because the utility of music usage data can be increased only when there is continuity of the same music. Figure 6 shows an example of a polymorphic pseudonymization algorithm module that makes the de-identified data stable. The advantages of the polymorphic pseudonymization process are its competitiveness, wide-spread data usage, and independence of the data. Therefore, it could be helpful for effective human resource and service improvement.



**Figure 5.** A process diagram for managing the unique key of an independent organization.



**Figure 6.** An example of a polymorphic pseudonymization algorithm module.

#### 4.2. Applications Using Music Usage Data

In this section, applications using music usage data will be proposed. One serves to detect an act of hoarding music, and the other serves to predict the marketability of music.

##### 4.2.1. Detecting an Act of Hoarding

There are various applications involving pseudonymized music usage data. One can expect that such a system could be applied to detect an act of hoarding, because outliers in music usage records could be detected. The so-called act of hoarding refers to an illegal act of manipulating music charts and real-time streaming rankings so that the songs of a specific singer are ranked. Several studies have been conducted to detect fraud and outliers in music streaming data [13,14]; however, the detection of an act of hoarding with limited music usage data has rarely been researched. An AutoEncoder model can be developed for detecting abnormal music usage records using de-identified music usage data [15]. This model discriminates the noise from the results, rendering the input data and output data equal.

We conducted an experiment with around 4 million songs' usage data from Genie Music. These data included the number of listeners, the streaming count for each hour, and the accumulated number of listeners and streaming count. We added an index to detect fake streaming, the fake-stream-count, which discriminates fake streaming counts from real streaming. The added index (fake-stream-count) can be calculated using the following equation:

$$(\text{fake-stream-count}) = (\text{play-in-term}/\text{person-max-stream-in-term}) - (\text{listen-in-term}), \quad (2)$$



where (play-in-term) is the number of total streaming counts, (person-max-stream-in-term) is the maximum number of streaming counts for one person, and (listen-in-term) is the number of listeners per hour. From this index, we developed the proposed AutoEncoder model through the normalization of the index. Figure 7 shows some sample songs taken from among 4 million songs' usage data from Genie Music with several indexes. We tested 3,945,728 data points from December of 2019 to October of 2021. Among the sample data, there were several anomaly points that were dramatically increased, as shown in Figure 8. We took the softmax function as an activation function for minimizing the loss rate. The epochs included approximately 100 tests; the batch size tested was 64~256; and we tested 10, 25, and 50 layers for the final model. Regarding the unsupervised deep learning AutoEncoder model, there was no final goal or objective, so it could be used to analyze several negative values of (fake-stream-counts) as anomaly points, even though these were normal data. According to the results, we could decide if the number of people who streamed the song was smaller than the (listen-in-term) number, meaning that one person listened to a specific song repeatedly [16]. In this case, we could identify one of the candidate songs as subject to an act of hoarding. Before the proper test using the log item standard field data shown in Table 1 was conducted, our preliminary model was tested using a fragmented music usage dataset. We performed the F1-score calculation by using data. The data were randomly chosen from the timepoints of 11:00 am, 8 December 2019 to 12:00 pm, 9 October 2020. During that period, a total of 6236 songs, excluding duplicates, were tested for our proposed model, and 29 songs were derived as anomalies. We set the criteria for F1-score evaluation, and the result was 0.95 [16]. This evaluation could be well derived using the log item standard field. To the best of our knowledge, the proposed method would be the first to be used for music usage data in Korea.

Subject	Singer	Rank_out_term	Play_in_term	Listen_in_term	Play_total	Listen_total	Person_max_stream	Count_person_stream	Fake_stream_count
BLUE	Sungwoon Ha	2,322	1,826,341.0	13,778.0	11,697,817	297,853	20	91,317.05	77,539.05
BIRD	Sungwoon Ha	2,513	1,986,610.0	14,350.0	22,987,617	491,911	20	99,330.50	84,980.50
Snowman	Seunghwan Jeong	5,790	3,411,068.0	96,025.0	47,818,879	2,133,539	20	170,553.40	74,528.40
180degree	Ben	6,627	3,463,172.0	91,301.0	54,261,232	2,180,574	20	173,158.60	81,857.60
I am the best	2NE1	8,779	1,265,986.0	155,784.0	5,540,948	777,700	20	63,299.30	-92,484.70
Good old days	Deokcheol Jang	4,254	2,657,639.0	57,417.0	89,705,216	2,751,023	20	132,881.95	75,464.95
Who U Are	Daniel Kang	1,992	1,804,754.0	13,736.0	8,978,586	202,415	20	90,237.70	76,501.70
My Way	ISU (MCtheMAX)	8,772	5,119,215.0	176,123.0	54,742,005	2,282,870	20	255,960.75	79,837.75

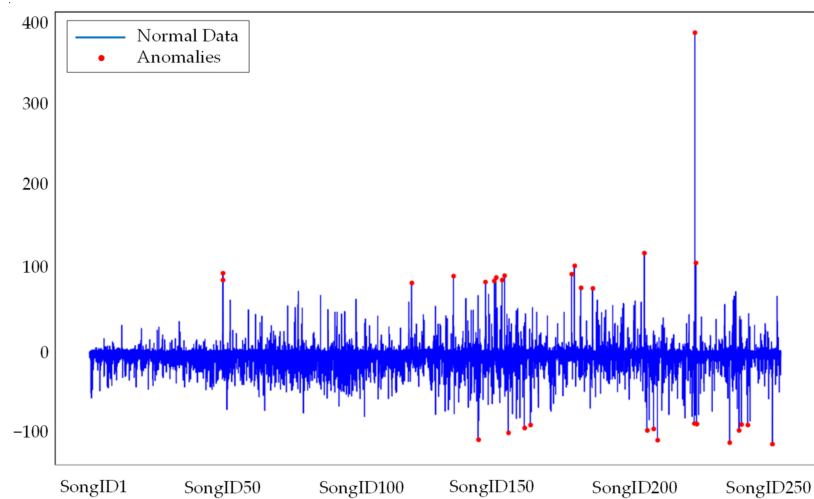
**Figure 7.** Sample songs, including usage data with several indexes (several songs and singers were written in Korean).

#### 4.2.2. Prediction of the Marketability of Music

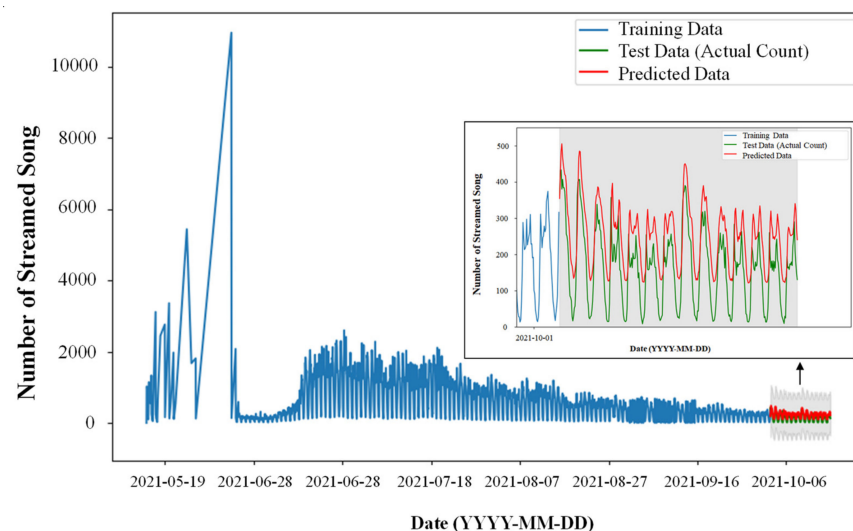
Another utilization of music usage data is to predict the marketability of music. Previously, the Korean music industry was limited to a focus on music sales. However, as the music industry is a fairly large market, it could be expanded to other fields. For example, by using music usage data from the past, it may be possible to predict the marketability of specific music or artists, and this could help to expand the scope of the industry through predictions based on data, as in the case of Moneyball in baseball or Opta in soccer.

Music usage data are time-series numerical data and are non-stationary; thus, we used the ARIMA model for the prediction of the marketability of music. ARIMA means AutoRegressive Integrated Moving Average, which is effective for analyzing time-series data [17]. However, music usage data are non-stationary, so we altered our data to make them appear stationary. The prediction data could be generated using the ARIMA model from the Python library, and the number of listeners per hour, shown in Figure 7, was needed for the prediction. For the proper ARIMA model, we adopted the AIC (Akaike Information Criterion) score, which is a criterion used to evaluate the quality of a statistical model. Using this, we set boundaries of error and improved the accuracy of the prediction

in terms of the distance between the prediction data and the actual data. Figure 9 shows the validity of our proposed model during the five months after the release of a specific song. The blue line indicates the number of streaming counts of the song as training data. The green line and red line show the difference between the actual streaming data of the song and its predicted data. We can see a similar tendency after October 2021. As shown in Figure 9, the streaming number of the streamed song shows a sharp increment at the release time and then stabilizes. We concluded that our model is only effective when it is independent of external valuables. For example, when the song was disseminated by the mass media or news, the streaming number could increase. Our model could not reflect this; however, it could be supplemented with an index of SNS data or news crawling data as input data. One can expect that SNS data such as the Google trend value for the specific song could be an index as an active valuable. An evaluation metric such as MAPE could be a candidate for estimating accuracy; however, the precision could vary depending on various conditions. The results could be enhanced through the elaborate adjustment of the index or by combining various indexes.



**Figure 8.** An example of an anomaly in the detection of music streaming. The horizontal axis denotes each song, and the y axis denotes the normalization results of the AutoEncoder model. A red dot indicates anomalous data of the streamed song.



**Figure 9.** A prediction graph based on music usage during five months (19 May 2021 to 6 October 2021) after releasing a song. The blue line indicates training data, the green line indicates test data (actual data), and the red line indicates the predicted data after five months.

## 5. Legislation Plan

As mentioned in Section 2, the current log collection system collects log data through the MOU, and there is no compulsion issue. The structural limitations of the current music data's usage are due to the structure of the music streaming platform and the fragmentation of the music copyrights of individuals. For most music streaming platforms, the royalties are paid according to the ratio by summing up all the payments made by all users and dividing them by the streaming number per song. This enables one to manipulate the sales volume or the music charts using a few accounts, and this may even distort royalty distribution. The fragmentation of the music copyrights of individuals can be solved by the individuals themselves, because they can only request data regarding how many times the song has been streamed.

Since the identification of several embezzlement issues, the Ministry of Culture, Sports, and Tourism has publicized the integrated music network and announced that they will revise and improve the system. However, the amended Copyright Act did not have direct clauses on music usage information collection; it is not appropriate that the operating plans for log collection systems are prepared based on the Copyright Act for the current legal system. However, there is room for adequate law revision in the following areas, because there are many provisions in the Music Industry Promotion Act:

- Provision location (provision);
- Integrated log collection system operator and duty (Minister of Culture, Sports, and Tourism);
- Obligation to join the log collection system (compulsion);
- Obligation to transmit usage information (technical issues and detailed adjustments as necessary);
- Establishment of support grounds;
- Other aspects (the scope of obligated subscribers, the scope and cycle of transmission data, and methods).

However, there are several obstacles involved in this legislation. First, there could be a procedural burden [18,19]. Many steps will be required in terms of the procedure, and such legislation cannot be developed unless there is someone to play a leading role. Second, there could be resistance from the obligated party, which is the music distributor. For example, the amendment to the Music Industry Promotion Act that intended to impose an obligation on music record producers to record performance information was proposed to the National Assembly; however, it ended up failing due to strong resistance from music record/music video production business operators. Third, coordination among the stakeholders is required. Various stakeholders exist due to the nature of the log collection system, and initiating a government-led system without gathering their opinions could cause many problems or even result in rejection of participating in the worst-case scenario. Finally, since this system needs to be institutionalized and organized, it will involve taxations, possibly resulting in public opinion backlash. Therefore, legislation should be developed in consideration of these obstacles. One might compare this situation to a black box, where there is only a problem instead of a settlement.

## 6. Conclusions

This study discussed the limitations and directions for the development of the music log collection system for the transparent settlement and distribution of music royalties. It first analyzed the log collection status and structural limitations of the current music usage information in Korea and suggested a transparent settlement and distribution structure for music usage based on the analysis. The results of the constructed pilot module were derived through the proposed settlement and distribution structure, and the music usage data were pseudonymized to examine several applications using AI models. Finally, we described ways to legislate this structure as well as the procedural difficulties of such legislation.

Unlike the DDEX structure, our proposal is for a settlement–distribution module that can be implemented based on limited music usage data. Therefore, there is a limitation

that music rights holders register their own music voluntarily in the proposed system. Additionally, the music log usage data are collected with the cooperation of major OSPs, meaning that eventually, global operators will also have to be included in the proposed system. We also contributed to research on AI applications based on music usage data. We believe that these contributions could enable an improved prediction of marketability if additional information on music usage data were to be provided. The current log collection system in Korea will hopefully be used for public purposes in a more progressive way.

**Author Contributions:** Conceptualization, Y.K. (Youngmin Kim), D.K. and S.P.; funding acquisition, Y.K. (Youngmin Kim), Y.K. (Yonghwa Kim), J.H., S.H., J.J., B.L. and H.O.; investigation and methodology, Y.K. (Youngmin Kim), D.K., S.P. and Y.K. (Yonghwa Kim); project administration, Y.K. (Youngmin Kim), resources, D.K. and S.P.; supervision, Y.K. (Youngmin Kim); writing of the original draft, Y.K. (Youngmin Kim), D.K. and S.P.; writing of the review and editing, Y.K. (Youngmin Kim) and Y.K. (Yonghwa Kim); validation, D.K., S.P. and Y.K. (Youngmin Kim); formal analysis, Y.K. (Youngmin Kim), D.K., S.P. and Y.K. (Yonghwa Kim); data curation, S.P. and Y.K. (Youngmin Kim); visualization, Y.K. (Youngmin Kim), D.K. and S.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports, and Tourism in 2022 (project: development of a copyright information verification platform utilizing massive sound source log information data; project number: CR202104005; contribution rate: 100%).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Anderson, C. *The Long Tail: Why the Future of Business Is Selling Less of More*; Hyperion: New York, NY, USA, 2006.
2. Easley, D.; Kleinberg, J. Power Laws and Rich-Get-Richer Phenomena. In *Networks, Crowds, and Markets: Reasoning about Highly Connected World*, 1st ed.; Cambridge University Press: Cambridge, UK, 2010.
3. Salganik, M.J.; Dodds, P.S.; Watts, D.J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **2006**, *311*, 854–856. [CrossRef]
4. Parc, J.; Kim, S.D. The Digital Transformation of the Korean Music Industry and the Global Emergence of K-Pop. *Sustainability* **2020**, *12*, 7790. [CrossRef]
5. Copyright Law in Korea. Available online: <http://copyright.or.kr/eng/laws-and-treaties/copyright-law/act.do> (accessed on 31 July 2023).
6. DDEX. Available online: <https://ddex.net> (accessed on 31 July 2023).
7. Distribution Rules in Korea (Amended 20 September 2019). Available online: <http://komca.or.kr/foreign2/eng2/K0403.jsp> (accessed on 31 July 2023).
8. Passman, D.S. *All You Need to Know About the Music Business*, 10th ed.; Simon and Schuster: New York, NY, USA, 2019.
9. Personal Information Protection Act in Korea. Available online: [https://elaw.klri.re.kr/kor\\_service/lawView.do?hseq=53044&lang=ENG](https://elaw.klri.re.kr/kor_service/lawView.do?hseq=53044&lang=ENG) (accessed on 31 July 2023).
10. Elton, S. Standardized Identifiers and Electronic Data Interchange Rules for Music (29 June 2021). Available online: <https://ssrn.com/abstract=3876748> (accessed on 31 July 2023).
11. Aamot, H.; Kohl, C.D.; Richter, D.; Knaup-Gregori, P. Pseudonymization of Patient Identifiers for Translational Research. *BMC Med. Inform. Decis. Mak.* **2013**, *13*, 75. [CrossRef] [PubMed]
12. Ribeiro, S.L.; Nakamura, E.T. Privacy Protection with Pseudonymization and Anonymization in a Health IoT System: Results from OCARIoT. In Proceedings of the IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019.
13. Mucic, N.; Filzmoser, P. A multivariate test for detecting fraud based on Benford's law, with application to music streaming data. *Stat. Methods Appl.* **2021**, *30*, 819–840. [CrossRef]
14. Herskind Sej, J.; Christiansen, T.; Dvinge, N.; Hougesen, D.; Schneider-Kamp, P.; Zimek, A. Outlier detection with explanations on music streaming data: A case study with danmark music group ltd. *Appl. Sci.* **2021**, *11*, 2270. [CrossRef]
15. Zhou, C.; Paffenroth, R.C. Anomaly Detection with Robust Deep Autoencoders. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017.

16. Chorocksoft. CIVP-AI, Github Repository. Available online: [https://github.com/chorocksoft-git/CIVP-AI\\_anomaly\\_detection](https://github.com/chorocksoft-git/CIVP-AI_anomaly_detection) (accessed on 31 July 2023).
17. Ho, S.L.; Xie, M. The Use of ARIMA Models for Reliability Forecasting and Analysis. *Comput. Ind. Eng.* **1999**, *35*, 213–216. [CrossRef]
18. The 2 Copyrights in a Song (and How to Make More Money from Them). Available online: <https://lawyerdrummer.com/2020/10/the-2-copyrights-in-a-song/> (accessed on 31 July 2023).
19. Dicola, P.; Touve, D. Licensing in the Shadow of Copyright. *Stanf. Technol. Law Rev.* **2014**, *17*, 397.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.