DIGITAL
MEDIA
FINLAND

Finnish Music Publishers' Association
Finnish Musicians' Union
Finnish Society of Composers and Lyricists (Elvis)
The Society of Finnish Composers
Teosto

Metadata of Digital Music Files: Summary

25.9.2017
Jari Muikku

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Background

The exponential growth of the individual plays of tracks on streaming services has brought along new kind of challenges for licensing, reporting and distribution. It is of paramount importance that the systems function seamlessly and automatically in order to achieve efficiency. However, the current process of producing and delivering metadata from one party to another is still far from perfect.

From the right holders' point of view the current situation causes two kinds of problems. Firstly, streaming services and parties involved in distributing the royalties have problems with matching the data and the money. Secondly, the moral rights do not actualise, as the names of the right holders are not made available to the end users of the streaming services, the consumers.

1.2. Aims

The aim of the study was to produce information on the current status of the metadata delivery chain and to make suggestions how to correct or improve the flow of metadata. The focus of the study was on the Finnish market situation.

The study deals only with new and future releases, as the ways to improve the quality of the metadata of the back catalogue was out of scope. Furthermore, the study concerns only audio files and not any kind of music-related audio-visual files.
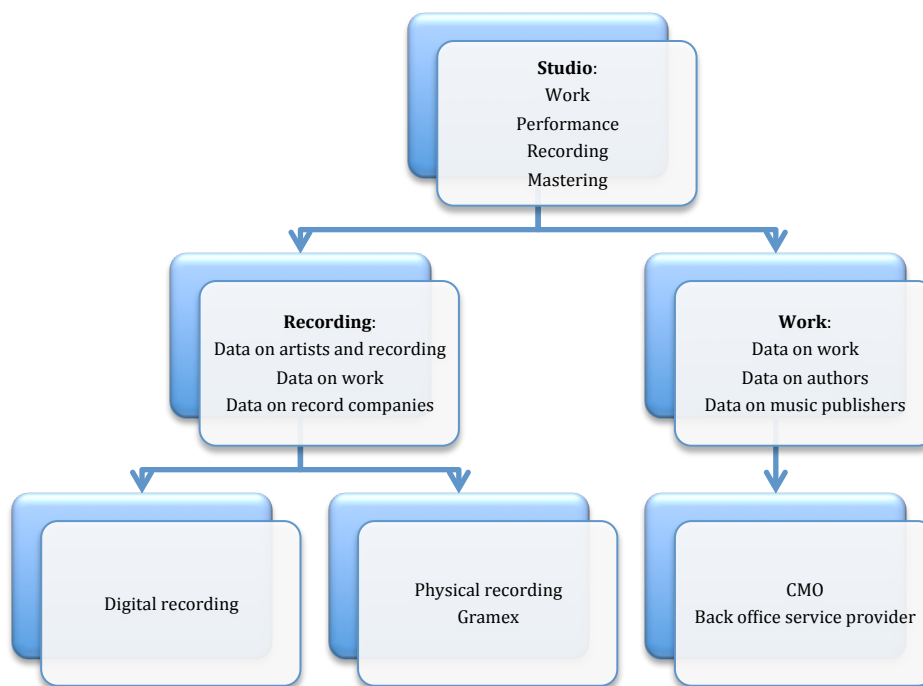
1.3. Methods

Consultant Jari Muikku of Digital Media Finland did the study during the period of April-June 2017. The project had a steering group, which consisted of the representatives of the institutions who commissioned the study. The work was supported financially by the Ministry of Education and Culture.

The study is based on two main sources, literature and interviews. Muikku did 24 interviews with total of 32 persons. These interviews covered all interest groups involved in the metadata process from individual composers and musicians to streaming services. The report is written so that confidential information was not disclosed.

2. PRODUCTION AND DELIVERY OF METADATA

2.1. Delivery Chain

The first step of the delivery chain is the performance and recording of a work in a studio. After this the recording is mastered for release in various forms, both physical and digital.

```
┌─────────────────────────┐
│      Studio:            │
│       Work              │
│    Performance          │
│     Recording           │
│     Mastering           │
└─────────────────────────┘
        │              │
┌───────────────────┐  ┌───────────────────────┐
│   Recording:      │  │      Work:            │
│ Data on artists   │  │  Data on work         │
│ and recording     │  │  Data on authors      │
│ Data on work      │  │  Data on music        │
│ Data on record    │  │  publishers           │
│ companies         │  │                       │
└───────────────────┘  └───────────────────────┘
    │          │               │
┌─────────┐ ┌──────────────┐ ┌──────────────────┐
│ Digital │ │ Physical     │ │ CMO              │
│recording│ │ recording    │ │ Back office      │
│         │ │ Gramex       │ │ service provider │
└─────────┘ └──────────────┘ └──────────────────┘
```

Picture 1: The flow of metadata from studio to databases

The new works are registered in a CMO (collective management organisation) who makes the information available to other relevant parties. The responsible studio producer and the A&R department of the record company usually collect the metadata concerning a recording. In case of a licensing or distribution deal the production company or another label is involved as well.

In Finland studio professionals used to collect recording data into the so-called studio lists but nowadays only few use them any more. The data is collected nowadays mainly by making notes and confirming them by e-mails or similar kind of electronic messages.

As the main aim of the studio work is to create the best possible atmosphere for the best possible performances, producers do not always give their full attention to administrative tasks.

In case of small indie labels the activities are often more hobby-like than professional by nature, and they might lack skills and knowledge on the music industry and its practises. For example, all are not even aware of the ISRC code.

The most common challenges and problems of the metadata process in the studio phase are:

- Mistakes in registering all musicians and their parts, and general communication breakdowns or misunderstandings
- Producers do not have financial incentives to produce and deliver top quality metadata unless they are acting also as co-writers or musicians
- Musicians and authors do not always check or do not have a chance to check data concerning themselves and their works

- In case of a new work the delay between commercial release and registering it into CMOs' or back office services' databases
- The changes of work titles after recording it
- The confusions between resembling work names
- The problems with matching CMO registrations if a work has more than one author and publisher
- Misspellings and problems with transliterations (e.g. Russian classical composer names) and Scandinavian special vowels

All in all, these problems are, according to the interviews, more occasional than systematic.

The role of music publishers varies a lot depending on a case. Quite often publishers are not at all involved in creating metadata. Record companies do not usually add publisher information to metadata unless it is easily available.

When the master files are ready, they are delivered to record companies. After this there are two different tracks for the metadata:

1. *Physical releases:* The master file and its metadata include usually all data on musicians and basic information on works (so-called label copy data) as the information is delivered to and used by Gramex, the copyright society for neighbouring rights.
2. *Digital releases:* There are two further tracks:
   o <u>Major record companies</u> deliver masters and metadata directly to DSPs (digital service providers)
   o <u>Indie companies</u> usually deliver masters and metadata first to distributor (either major label or aggregator), who delivers them further to DSPs
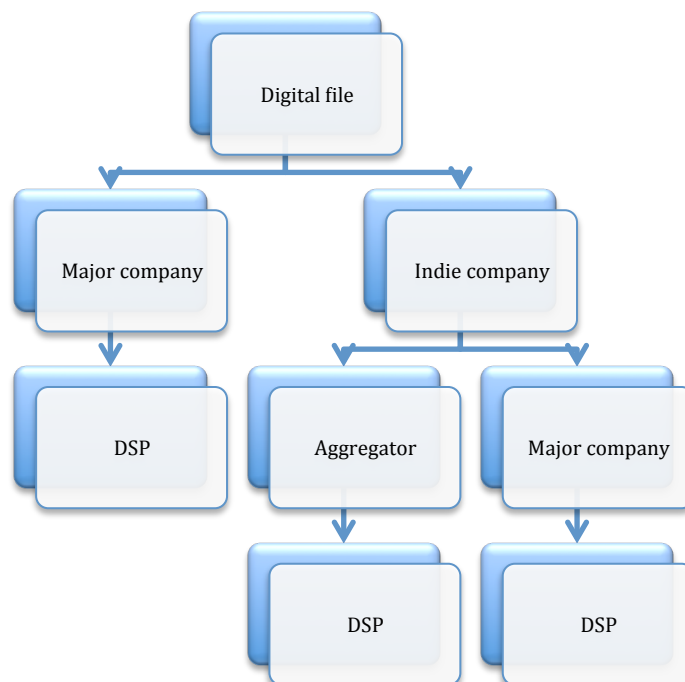
Picture 2: The flow of metadata from record labels to DSPs

The metadata processes of the three major record companies are similar among themselves, and they are same both to their own products and the products they license or distribute.

A&R department collects the metadata and passes it on to the employees who are responsible for the product administration. They feed the data into the companies' internal data systems a couple of weeks before the release date. Each company's data system is global and centralised.

Unless all obligatory data fields are filled out, it is not possible to create a new product into the database. There are also optional fields available for additional information such as other musicians than the main artist(s).

The optional fields are not usually used unless they have additional value compared with the time and resources used for collecting and typing in the data, or if the data is not reliable. Finding publisher information is usually the hardest task. Furthermore, major labels consider that it is not their duty to add this information. However, in most cases the label copy information is available.

The obligatory information fields are usually:

- The name of the artist: main artist and featuring artist
- Title of the work
- Authors: composer(s), lyricist(s) and arranger(s) (optional field in one major company)
- Language (of the country of origin and used in the work)
- Duration
- Genre
- The country of the copyright owner
- Copyrights
- ISRC code
- UPC/EAN code(s)
- Company's internal code (either its own or based on GRid)

Optional fields are usually:

- Other musicians
- Publisher(s)
- Miscellaneous information

The systems of the major labels produce automatically the most important codes such as ISRC, EAN, and UPC. In addition to this, each company has its own codes for products and/or artists. None of the major labels use ISWC.

At the end of the process the centralised unit, which is based either in UK or US, checks out the metadata. After this the files and the respective metadata are sent to the DSPs.

The processes of the major labels are quite fluent and efficient. However, there were some challenges and problems:

- Mistakes in manual work phases such as typos
- Imprecise information: for example, composers or artists with a family name only can be mixed with another one
- No information on the artists' real name; a pseudonym can include several individuals
- Mistakes in back catalogue information for various reasons such as previous database mergers during acquisition processes
- Misinterpretations and misunderstandings
- Lack of original name of a cover song
- It is difficult to change information, which has been fed into the systems

The processes of indie companies are similar to major companies' but the quality varies a lot. This depends on the level of professionalism of an indie label. However, if an indie label works with a major label or a good aggregator their processes force to work in certain ways, and this reduces problems with metadata.

Some of the problems and challenges faced specifically by indie labels are:

- It might be difficult to find out what happens with the products whose distribution deals are expired
- Long physical distance with distributor may cause communication problems
- Uncertainty concerning the quality of information received from distributors

The processes of aggregators/distributors are similar compared to the major companies. Their clients have to fill in obligatory metadata fields before their products can be delivered to DSPs.

The aggregators have to modify the metadata for each DSP according to their individual style sheets. The delivery of music files and metadata between aggregators and DSPs is done according to the DDEX standards.

Some of the challenges and problems of the aggregators/distributors are:

- Indie company can fill in the obligatory fields but the quality of the data can be poor
- Small indie labels can have attitude problems with administrational processes and cause extra work
- All in all, the share of metadata problem cases is quite low
- There are some small distributors in business who allow and deliver metadata of poor quality
- The karaoke versions can cause problems with original works, and therefore some distributors reserve the right to erase them from their catalogues

DSPs receive music files and metadata according to DDEX standards. It was noted that DDEX is only a standard, which does not work without proper processes.

One of the main problems of metadata for DSPs is poor quality. Data can be formally ok but the content can be rubbish. DSPs have no means to control this, as they are totally dependent on the metadata delivered by major labels and aggregators.

Some DSPs have tried to improve the quality with incentives such as ranking systems, which give privileges for the parties who deliver metadata of good quality. At least one major DSP has author information as an optional field but it rewards parties who deliver this as well. Missing information will give lower rankings.

The most important codes for DSPs are ISRC and EAN/UPC. The usage of ISWC is problematic as they have not direct access to CMO databases. ISWC has also some challenging features such as the relation of an entire work and piece of it (for example, single movement of an entire symphony). DSPs are currently studying possibilities to use ISWC on a wider scale.
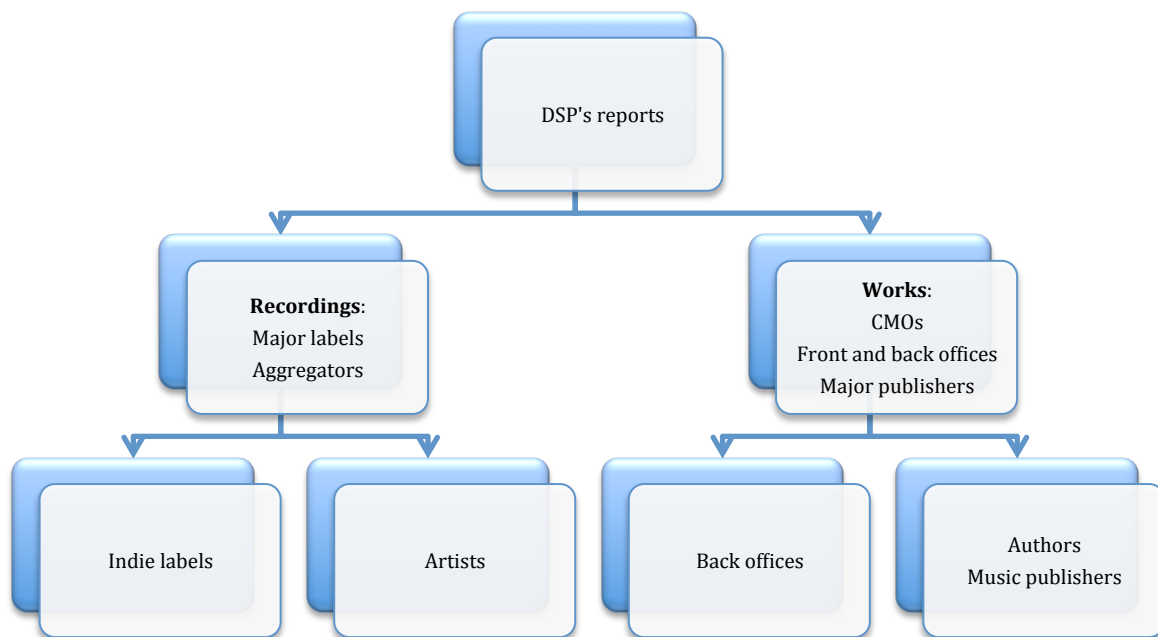
The problems mentioned earlier affect also DSPs. These problems recur in usage reports to right owners. Furthermore, DSPs have to filter information so that competitors do not get information on their shares of individual works or tracks.

There is a clear, at least 80:20 ratios between the economic value and the manual work. The Top 100 works and tracks are usually almost free of problems as all parties pay their full attention to them. However, in the Top 1000 already every tenth track lacks composer/author information. The need for manual work usually concerns works and tracks with less usage, metadata of poor quality, and which come from smaller distributors. The overall situation has improved gradually to certain extent as the processes have been established among all parties.

Making the names of authors and musicians available in the user interfaces or UXs of DSPs depends on the problems and challenges discussed above. If DSPs do not get the data from major companies or aggregators, or if the quality of the data varies too much, showing it in their UXs is not possible. It is particularly difficult to get information on individual musicians.

Another point with moral rights is the common practises. As the data has not been made available so far the DSPs do not consider it to be their duty. The longer the status remains the same the more passive the parties get.

The opinions of showing richer metadata to consumers are divided. On the other hand, major companies see that an average consumer could not care less about the additional information, and therefore it is not worth investing in. On the other hand, richer metadata was considered to serve active music fans and improve their commitment to the service, and encourage them to use more money. In addition to this, e.g. fans of classical music are used to get richer data, and jazz fans are interested in individual musicians.

```
                    ┌─────────────────┐
                    │  DSP's reports  │
                    └────────┬────────┘
              ┌──────────────┴──────────────┐
       ┌──────┴───────┐              ┌───────┴──────────────┐
       │ Recordings:  │              │ Works:               │
       │ Major labels │              │ CMOs                 │
       │ Aggregators  │              │ Front and back       │
       │              │              │ offices              │
       │              │              │ Major publishers     │
       └──────┬───────┘              └───────┬──────────────┘
        ┌─────┴─────┐                 ┌──────┴──────────┐
  ┌─────┴────┐ ┌────┴────┐     ┌──────┴─────┐ ┌─────────┴─────────┐
  │  Indie   │ │ Artists │     │   Back     │ │ Authors           │
  │  labels  │ │         │     │   offices  │ │ Music publishers  │
  └──────────┘ └─────────┘     └────────────┘ └───────────────────┘
```

Picture 3: The flow or DSPs' usage reports

Major labels get their reports and payments directly from DSPs according to the DDEX standards, and indie label via their distributors. This part of the process goes quite smoothly as it is based entirely on ISRC, EAN/UPC and other codes.

The reporting concerning the works is a more complicated process. DSPs have to filter the information according to the parties who represent various parts of the rights in various countries at a certain time. DSPs want to avoid getting multiple and overlapping invoices concerning the same rights from several parties.

The biggest challenge of the handling of the usage reports is to combine recording and work data, and to find the correct right holders. Problems occur, for example, in situations where different DSPs send different kind of information concerning the same track. Matching is usually done by combining various kinds of databases.

Back office services and CMOs have developed algorithms, which learn to correct obvious mistakes such as typos in common names. Services cannot afford doing manual matching work unless the case has big enough economic value.

The automatic handling of reports has improved during the past couple of years. For example Network of Music Partners (NMP) says that the index measuring the amount of data and its financial value is already close to 100. According to the NMP statistics, in June 2015 as much as 42% of Top 100 tracks missed information on works and publishers, and in January 2017 only 4 %. However, the situation is still far from good. At the top 10% of reported tracks, which represent more than 90% of commercial value, composer/author data is still missing from more than 1/3 of the reported tracks.

All in all, the situation has not improved as quickly as the parties had expected. One of the main reasons for this are the processes, which have remained more or less the same during the past decade or so. The usage of identifier codes and processes vary too much according to a party. The interpretations of contracts vary as well and cause challenges for efficient activities.

The biggest problems and challenges of handling DSPs' reports are:

- The complex ownerships and administration of rights of works
- The reporting is entirely based on the metadata produced by the record companies
- Combining various databases in order to improve auto matching
- Different parties send different information concerning the same track
- Limited possibilities for manual work
- Reports with works which have not been filed into CMOs' databases

2.2.   Some General Problems and Challenges

The flow of metadata is shattered and modular by nature. Each party sees and is concerned only of its own part of the process and not of the process as a whole. There is too little discussion on this among the industry parties. Each party has either adapted to the current situation or consider the problems to be beyond their control. The latter one concerns especially parties coming from small market areas such as Finland.

The national and international metadata development projects treat symptoms and do not affect the diseases or the root problems. Processes can be made more fluent but deep structures and operational cultures are difficult to change.

Even though music business has gone digital, the business culture is still rooted, at least partly, into the world of physical products.

3. SOLUTIONS AND SUGGESTIONS FOR ACTION POINTS

The following list deals in more detail the problems and challenges presented in the previous chapter. The order does not refer to the importance of the issues. These issues have been analysed mainly from the Finnish point of view.

The difficulty of executing each action point is estimated in scale 1 to 5 (1 = easy, 5 = very difficult). The short and long timespan effects are estimated in scale 1 to 5 as well (1 = minor, 5 = siginificant). Short timespan is 1-2 years and long 3-5 years.

1. The collection of metadata in studios is not systematic. Responsible persons (mainly studio producers) do not have financial incentives to produce data of best quality.

Possible solutions: Taking studio lists back in use in digital form. Creating incentives for producers to deliver metadata of best possible quality.
Action points: Market study for applicable products (such as Swedish Auddly or suitable blockchain-based products)
Difficulty: 3

Effects short/long timespan: 2/4
NB: Individual authors and musicians should be urged to check their own data

2. A new work may be registered to CMO databases only long after the release of its recording.

Possible solutions: Speedier registration process incl. composers and publishers
Action points: Constant informing about the importance of the quick registration of works. Highlighting how important it is for efficiency and avoiding delays in distribution.
Difficulty: 1
Effects short/long timespan: 2/3

3. Problems in matching registrations if a work have more than one composer and publisher.

Possible solutions: Harmonization of processes and usage of identifier codes. Development of registration systems.
Action points: Direct customer feedback to CMOs and back office service providers, and participating international development initiatives.
Difficulty: 4
Effects short/long timespan: 3/4

4. Attitude problems and lack of business skills among smaller indie labels, which produce metadata of poor quality.

Possible solutions: Training and education in various forms. Suitable combinations of sanctions and incentives set by distributors.
Action points: Music business parties create information packages, which are easily available and are updated on regular basis. Seminars and workshops.
Difficulty: 1
Effects short/long timespan: 1/2
NB: As the market share of smallest labels is rather insignificant, the overall effects are quite limited from the point of view of the entire industry. This would reduce in any case the need for manual work.

5. DSPs are depended of the metadata produced by the record labels. Getting information on works directly from CMOs is difficult. DSPs do not currently get reliable and comprehensive data on musicians anywhere.

Possible solutions: CMOs and back office service providers make their databases available in a more accessible manner. Studying possibilities of getting more extensive information on performers by using in a similar way the databases of neighbouring rights societies.
Action points: Continuous dialogue between DSPs and CMOs/back office service providers about the usage of their databases.
Difficulty: 4
Effects short/long timespan: 2/4

6. Moral rights are not fulfilled, as the UXs of DSPs do not provide users with information on authors and musicians.

Possible solutions: DSPs require this data and explore the possibilities of using fore-mentioned databases. Recording contracts include clauses concerning delivering complete performer data to DSPs. The additional value of presenting richer metadata in the UXs is studied: for example, using metadata for better ad targeting, differentiation of free and premium services, reducing CHURN, and deepening of artist-fan relationships. Implementation of moral rights by judicial means.
Action points: Business parties seek out in co-operation economic incentives for producing richer metadata for consumers.
Difficulty: 5
Effects short/long timespan: 1/4
NB: The consumers should also be made aware of this issue. Furthermore, richer metadata is important for authors and musicians in terms of marketing their skills and getting new work opportunities.

7. The complex ownership structure of works and administration.

Possible solutions: Making the databases available on wider scale than today. Collecting data into larger pools. Harmonizing of the usage of identifier codes and processes among industry parties.
Action points: Targeted joint development projects among PROs, publishers and back office services.
Difficulty: 5
Effects short/long timespan: 2/5
NB: There are several international development projects going on.

4. CONCLUSIONS

The root reason behind many problems is the traditional division in business between recordings and works. Both sides have their established models, practises and cultures, which are difficult to change or even unify within a short timespan. In addition to this, the business cultures of physical and digital recordings are still not alike, and unifying this will take some time as well.

This setting can be seen, for example, in the variability in the usage of identifier codes and the quality of processes. The result reminds the classic whispering game, where the short message may have finally altered significantly when it has been passed trough the line from one person to another.

The most commonly proposed solution for metadata problems is one global database. However, the history has proved that initiatives like Global Repertoire Database have failed due to financial and political disagreements. The current prevailing approach is to combine various kinds of data pools by using open and standardised interfaces.

Another problem of metadata development projects is lack of depth. On the other hand, progress has made in the field of technical solutions, but, on the other hand, they have not touched the issues of underlying structural and cultural issues.

The common nominator for all initiatives is a motive, which can be endorsed by all business parties: finding big enough economic incentives, both in forms of additional income and savings.

Sanctions, which are based either in bi- or multilateral contracts or legislation, are problematic and difficult to implement due to the global nature of business and the challenges of enforcement.

However, in everyday business both sanctions and incentives are needed. The relation between "stick and carrot" should be pondered in each metadata development project. At the moment, all parties go for the same goal, growing the value of the streaming market. This should be taken as a key point in all metadata development projects.