



<http://www.diva-portal.org>

This is the published version of a paper presented at *eChallenges e-2009 Conference, Istanbul, Turkey, October 21-23, 2009*.

Citation for the original published paper:

Johansson, D. (2009)

An automated system for analyzing music usage and metadata exchange on digital music services

In: Paul Cunningham & Miriam Cunningham (ed.), *eChallenges e-2009 Conference Proceedings* (pp. 1-9). International Information Management Corporation Limited

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-72683>

An Automated System for Analyzing Music Usage and Metadata Exchange on Digital Music Services

Daniel JOHANSSON

Royal Institute of Technology, Media technology and Graphic Arts
Musiclink Hultsfred
Tel: +46 70 399 05 00, Email: daniel.johansson@musiclink.se

Abstract: The study investigated music usage on several online services over a period of sixteen months with the purpose of identifying listening behaviours among fans and differences between music genres online. The research was conducted with the Music Intelligence Mashup System, a platform for automatically collecting data from online services. The study identifies large discrepancies between genres as well as a number of interesting findings on listening behaviours for particular artists. The paper also presents several suggestions for how the music industry could develop new business models in the digital environment.

1. Introduction

As new technologies for digital distribution of music gets ever more important for record labels and other actors in the music industry there is an increasing need for new Business Intelligence tools. Music marketing and distribution in the digital environment is diversified and often decentralized, music flows through hundreds of services online, web radios, UGC services¹, music blogs, social networks, recommendation services, digital retailers and file sharing networks. [1] It is becoming more and more difficult to get an overview of how music is listened to online as the usage is moving from traditional media platforms like radio, television and tangible formats like the record or the magnetic tape towards digital music services like MySpace, YouTube, last.fm, iTunes, iMeem and Spotify. Music listening seems to have increased among certain groups in society during the latest years as a result of new digital technologies [2], but, at the same time it has become much more complex for actors in the music industry to follow how the popularity of a specific artist is developing over time. Before the Internet it was possible to measure specific artist's popularity by simply following record sales or concert ticket sales. Today it is much harder to see how, when and why popularity begins to grow as the audience themselves are much more involved in "breaking" the artist.

To understand the effects from online marketing and promotion campaigns, a music company nowadays needs to follow an increasing number of digital music outlets, something that is both time consuming and dependent on specific competence and knowledge that seldom exists at the music firm. In a former study of the Swedish music industry, the growing problem for music companies to get an overview of the development for specific artists online and create new business models based on this new knowledge, was highlighted: "The straggling approaches draw a diverse picture of a market where long-time strategies seem to

¹ User Generated Content

be hard to draw simply because there is not enough knowledge on how the market will evolve, not even on short-term basis. This together with mental models and beliefs about how the music business is supposed to work makes it hard for some record labels to adjust to the Internet economy and new consumer behaviour.” [3]

The above mentioned study showed that although there was a difference between how music companies were working online, there was also a consensus between the informants that they all needed better ways to collect and analyze music listening behaviours on digital music services. The need for automating this work instead of having to manually receive and handle statistics in different data formats, became obvious. In 2008 and 2009 a system was developed by Musiclink² to meet the needs from the music industry, the Music Intelligence Mashup System that automatically collects and analyzes data from a number of online services. In May 2009 the system was monitoring over 300 artists.

1.1 Literature Review

The paradigm shift happening in the music industry have been discussed and researched upon from different perspectives and disciplines. According to Théberge, technology has always been “a foundational element upon which the music industry is based, an inherent component of its operational strategies and its ultimate profitability.” [4] Hence, when the foundational element is going through rapid changes, so does also the organizational structures and business models.

Wikström found, after interviewing 36 stakeholders in the music industry, that record labels needs to find a new logic in their business models to meet the digital media environment: “The study argues that one of the most significant changes of the media environment is the development of new media outlets which are incompatible with the “traditional” media outlet structure. According to these traditional structures, a certain set of media outlets (primarily radio and television) is used by music firms to *expose* music to audiences, while a distinctively different set of media outlets (physical carriers such as CDs and Compact Cassettes) is used for *distribution* purposes. New media outlets are not as easily defined as either tools for exposure or tools for distribution.” [5] Media services like MySpace and YouTube are functioning as *both* marketing and distribution platforms, hence understanding the audience behaviour on such services is becoming increasingly important when a large portion of music usage is made on them.

Kusek & Leonhard [6] argues that the music industry is moving from a product based industry to a service based industry, they envision a future where music is available in the same easy way as electricity or water. Fischer [7] describes how a new payment model would make it possible to integrate the illegal decentralized distribution of today into the legal monetary system, creating new incomes for rights holders in the music industry. Other researchers have also tried to identify how the structure of the music industry is changing as a result of the ongoing technology shift [8][9][10]. New digital music services are standing in the middle of these changes as the infrastructure for revenues is changing [11]. To understand these services is therefore becoming a crucial part of running a music business today.

2. Objectives

The main objectives of the study is to:

- Describe the Music Intelligence Mashup System, an automated system for collecting and analyzing data from several digital music services.

² Musiclink is an ICT company owned by one of the largest music festivals in Sweden, the Hultsfred festival. The company is conducting ICT research and development for the music industry.

- Present an analysis of behaviours among music listeners for specific artists by comparing the collected data with marketing efforts done by an artist or music company.
- Identify differences between genres on different digital services.

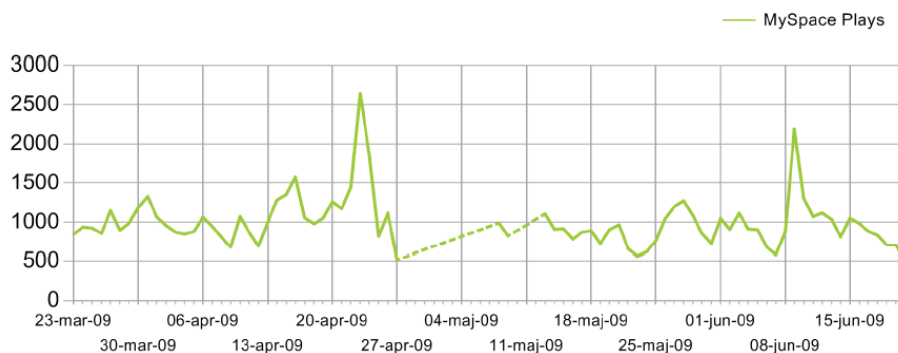
3. Methodology

The study was conducted in two phases:

- 1) System development and data collecting
- 2) Statistical analysis of the collected data

The system itself was developed on the .NET Framework with Workflow Foundation as core, using SQL Server as engine for the databases. The presentation layer uses SQL Server Reporting Services with Dundas Charts. An iterative system development methodology were used to build the Music Intelligence Mashup System (more details of the system is presented in 4.1) and statistical analysis were made on data from five digital services identified as important for music listening online: MySpace, YouTube, MySpace TV, last.fm and The Pirate Bay. In the cases of existing Application Programming Interfaces, these web services were used. In other cases web scraping technologies were used for gathering the data.

The study was conducted in cooperation with a number of companies from different parts of the music industry.³ The contribution from the actors included interviews with stakeholders to understand how the analysis of the collected data could be made in the most efficient way. The MIMS system collects data every hour from each service. By comparing the collected data with the existing figures in the database it is possible for the system to create a trend for the usage and plot a usage curve over time on the specific services. Below an example for the artist A Camp on the number of songs played daily on MySpace:



On MySpace the system automatically collects the number of times a profile page has been loaded, the number of times the songs have been played and the number of friends the specific profile is having. On YouTube and MySpace TV the system collects the number of times a specific video have been viewed. On last.fm the system collects data on the number of times mp3 files have been played on the registered users computers and on The Pirate Bay the system collects the number of seeders and leechers for a specific torrent. The system is then automatically adding the different data streams to a global development graph. During the study, reports were sent automatically every Monday to the partners in the project and feedback was provided from the companies making it possible to identify reasons for peaks or other changes in the popularity trend of the artist. Many of the partners provided internal data on promotion campaigns and releases that we could correlate with the data collected, making it possible to identify how large impact media appearances, concerts, record releases etc had for the online interest of the particular artist.

³ Artist Company Ten, Universal Music Group, Songs I Wish I Had Written, Swedish Performing Rights Society (STIM), Swedish Independent Music Producers (SOM) and Rockparty.

3.1 Limitations of the study

The study focused on Swedish artists and companies mainly. The part of the study where artists from different genres are compared with each other was limited to a total of 27 artists in four genres, these artists were chosen as most representative for each genre.

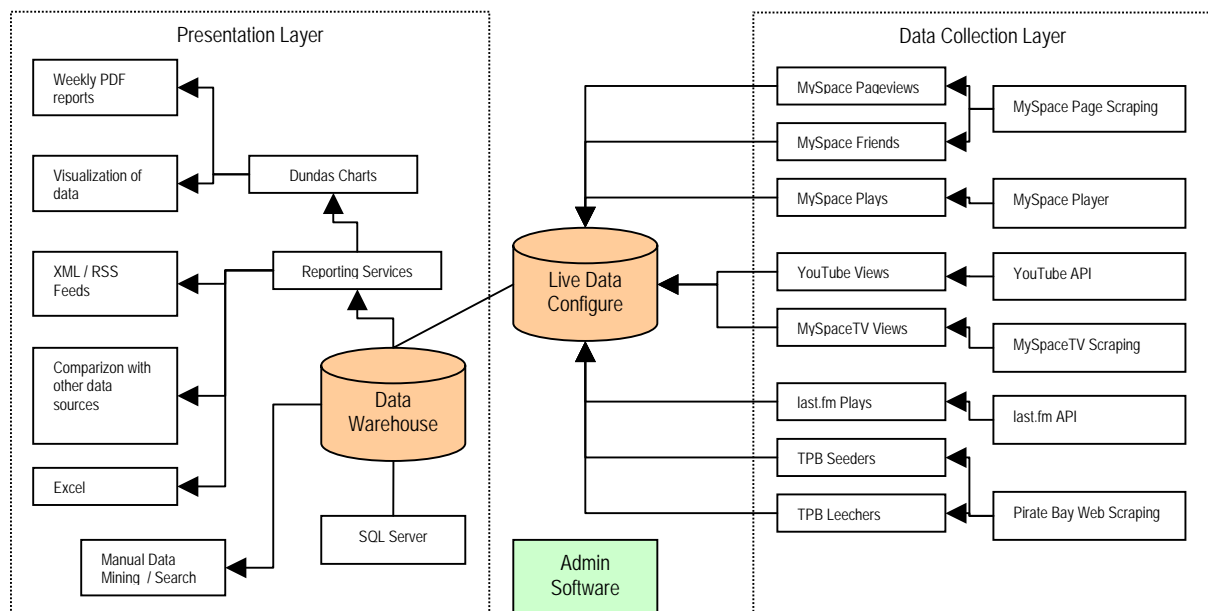
4. Findings

4.1 The Music Intelligence Mashup System Architecture

Data provided by the services are giving the music owner a hint on the popularity and interest of the music at the service, but when trying to understand the popularity over time it becomes somewhat difficult to follow the development. The Total Plays category on MySpace, showing that a song has been played 82.344 times for an artist does not say anything about when the plays were made. It might be possible that 75.000 of the plays were made one year ago and only 12.344 plays were made during the latest year leading to the conclusion that the interest of the music has dropped dramatically. In order to get a detailed understanding of the music usage over time it is therefore necessary to regularly collect data from the services and compare it to understand the trend. MIMS is automating this process.

The data collection system was built using web scraping scripts as well as Application Programming Interfaces in the cases where such were available from the online services. The system was developed on .NET Framework with Workflow Foundation as core, using SQL Server as engine for the databases. A number of collection modules were built running as Windows services on a dedicated server. These modules are governed by an administration software so that intervals and data sources can be changed according to data collection needs. The five services chosen were MySpace, YouTube, last.fm, MySpaceTV and The Pirate Bay (TPB). These services were chosen primarily because of their importance for the artists and genres in the study.

Architectural Overview

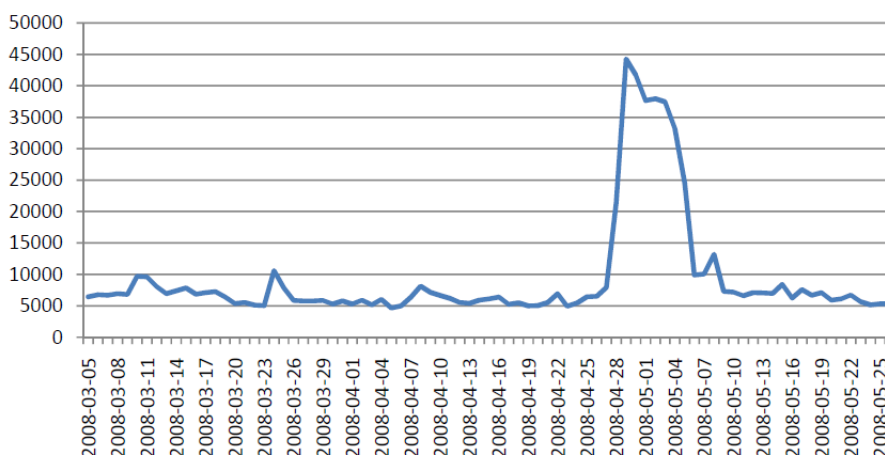


Although the separate data streams collected is not unique in itself, the automatic analysis and work up of the statistics creates an effective and simplified process for the client to follow the development of a specific artist at several digital music services. Instead of having to manually check each of the services own statistics, something that is both time consuming and

somewhat difficult for the music firm, the client receives a daily or weekly report over the monitored artist.

4.2 Listening Behaviour for Specific Artists

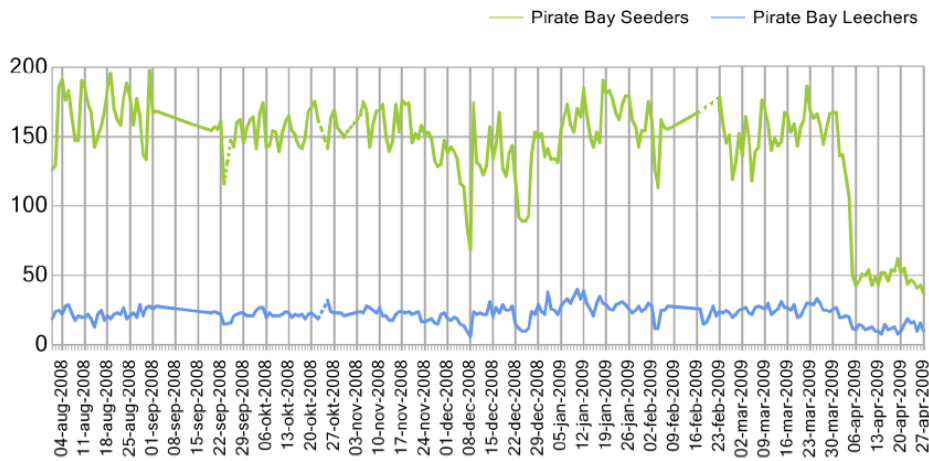
By following and analyzing individual artists we have been able to find several interesting behaviours and listening habits among fans for the 300 artists monitored in the study, findings that we believe are of great value for the artists and companies working with the artists, i.e. record labels, booking agencies and managements. For example, when the Swedish artist Robyn appeared at the Letterman show in United States in May 2008, there was a huge increase in the number of song plays at MySpace. What is even more interesting is the fact that the interest were at these high levels for almost a whole week:



The direct implications of such a finding is that if the music company can foresee such an increase in traffic, they can use it as a window of opportunity for selling ads, make specific campaigns, contests or in other ways use the increased interest. The same kind of increases can be seen on a number of other artists we monitored in the study (Erik Hassle, Moto Boy, Takida, Agnes etc.). By analyzing these figures it becomes possible for a music company to make a prognosis for online effects of media appearances and in a much more conscious way use the interest during short periods of time to increase revenues and guide traffic to the artist's web site.

We also found several important listening cycles depending on what kind of genre the artist's music is related to. The Swedish artist Familjen is producing music in the Dance/Electronica genre, and we found that Friday and Saturday evenings are the time periods when most fans are listening to the Familjen music on YouTube and MySpace. As the music could be categorized as somewhat "party music", the reason could simply be that people are listening to the music when they are having parties at home or preparing for an evening at the clubs. These periods of time could also be used in specific ways to connect to fans when they are in the "right mood" and create contests or campaigns that can be more efficient than regular ones that are online all the time.

Another finding was how the implementation of the EU Intellectual Property Rights Enforcement Directive in the Swedish copyright law influenced file sharing of Swedish artists on April 1, 2009, when file sharers obviously became afraid of sharing music illegally online. Familjen was one of the Swedish artists where this was most notable. Below a graph showing the numbers of seeders and leechers on all torrents related to the artist at The Pirate Bay:



By comparing data from different artists with each other it also becomes possible to find out how well an artist is doing online at particular services and understand the development of the online presence by creating a popularity “index” for the firm. For example, the Swedish artists Markus Krunegård and Erik Hassle had almost the same amount of song plays and video views at MySpace and YouTube when data was compared on April 13, 2009, Markus Krunegård had 908.599 video views at YouTube and 311.225 song plays at MySpace, and Erik Hassle had 808.797 video views at YouTube and 231.555 song plays at MySpace. But it becomes rather interesting when the number of song plays played through last.fm is compared, 863.552 song plays for Markus Krunegård but only 26.962 song plays for Erik Hassle. The core reason for this could be the fact that at the time of comparison Erik Hassle had not yet released his debut album, hence there was not so many songs out on the Internet that people could download and scrobble on their computers⁴. Still, although Erik Hassle had not yet released his album, the artist was almost at the same level as Markus Krunegård at both YouTube and MySpace. Furthermore, when data was compared over time, it became obvious that the trend for Markus Krunegård was the same on last.fm but were shrinking at MySpace, while for Erik Hassle the interest on both MySpace and YouTube were constantly rising. Hence, a logical conclusion would be that the interest for Erik Hassle online is larger than what it was for Markus Krunegård when he released his debut album, leading to the preliminary prognosis that Erik Hassle might become a bigger artist online than Markus Krunegård when the album is released, and sell more of the album, making it possible for the record label to risk more resources in the marketing campaign. When comparing Markus Krunegård with other artists it also becomes obvious that his fans are very active in scrobbling their mp3 listening. Either the fans as such are a group that are more dedicated music listeners, since users of last.fm tend to be more musically interested, or it is simply a result of the fact that this particular artists fans are mainly listening a lot to the music locally on their computers rather than at online streaming services.

The main output from the comparisons made with these and other artists is an increased knowledge that is crucial for the music company. Although only in the beginning we can see that the companies cooperating with us in this study have started to change their routines and habits when analyzing the online behaviour of the fans of the artist they are working with. The data provided by MIMS can thus function as Business Intelligence for the firm to create new business models and revenue streams.

⁴ Audioscrobbing is a technology introduced and used by last.fm that automatically collects ID3 metadata from mp3 files when a user is playing the song locally on his or her computer.

4.3 Differences between genres on music services

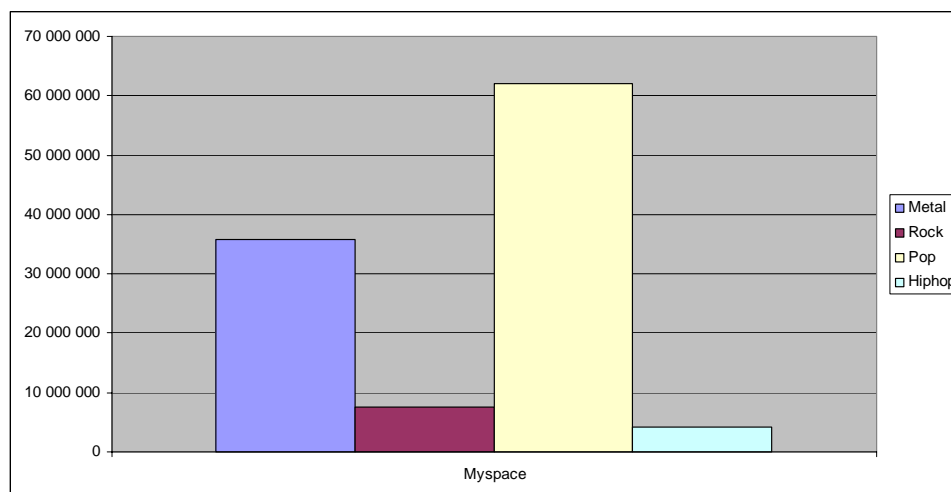
Based on the data collected with the Music Intelligence Mashup System we also compared large Swedish artists in different genres with the purpose of trying to identify differences or similarities between them. Four different genres were chosen: Pop, Rock, Metal and Hip-Hop, and artists were selected based on the bands own labelling of genre at MySpace. The selection of artists were made by analyzing top Swedish artists at MySpace top lists, sales charts at Hitlistan and a dialogue with music industry stakeholders. Also a general judgement based on knowledge of the particular artists by the researchers involved in the study were made.

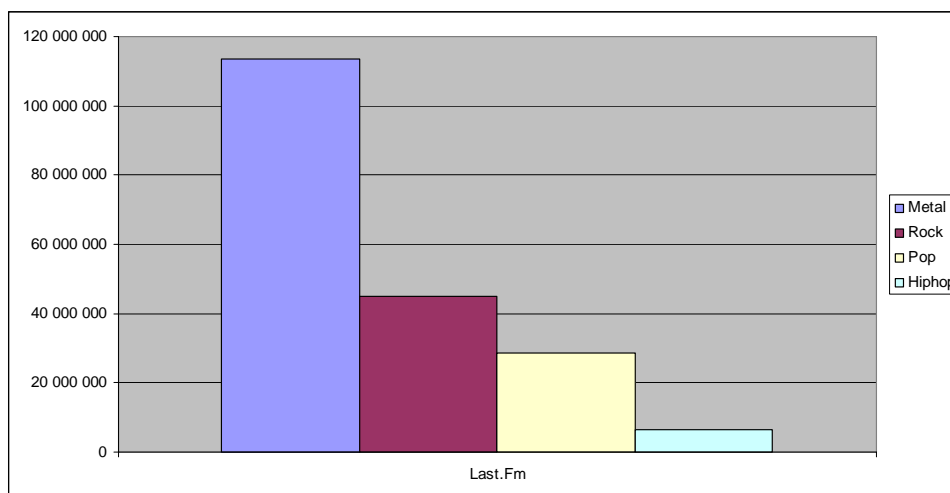
Data from the following artists were chosen: In the pop genre, Robyn, Mando Diao, Basshunter, September, Per Gessle, EMD, Måns Zelmerlöw and Amanda Jenssen, in the rock genre, Melody Club, The Hives, Kent, Lars Winnerbäck, The Ark and Brolle Jr, in the hip-hop genre, Json, Petter, Adam Tensta, Afasi & Filthy, Looptroop and Funky Fresh and in the metal genre, Sabaton, Hammerfall, In Flames, Meshuggah, The Haunted, At the Gates and Opeth.

By comparing data from the different services we were able to find that:

- Swedish Hip-Hop fans are not primarily interested in listening to artists on MySpace, but rather on YouTube.
- Metal fans are listening to music files locally at their computers to a much larger extent than in the other genres.
- The best platform for promoting music in the pop genre for Swedish artists is YouTube.
- Basshunter is by far the largest Swedish pop artist online during the time period. The second largest pop artist at YouTube, September, had just 20.51% of the amount of views that Basshunter had during the period.
- Although The Hives was a very popular band on MySpace during the time period, with over 5 million song plays, their popularity at YouTube did not correlate. They were just the third most viewed rock band at YouTube, while they had 711% more song plays on MySpace compared with the second largest rock band, Kent.
- In the metal genre, In Flames stands out as having a very strong online presence. They were the largest band on MySpace, YouTube and The Pirate Bay, and had more plays through last.fm than all the other metal bands together during the study.

The graphs below shows the differences between the MySpace and last.fm services when the genre totals are summed up:





5. Conclusions and Suggestions

Although some of the data collected from the different services in this study themselves are not unique, the approach made possible with the Music Intelligence Mashup System simplifies the work for a music company to get an overview of the development and popularity of their artists online. Furthermore, the 35 million data posts makes it possible to compare genres and artists at different services and find correlations as well as differences between listening behaviours and music usage. This knowledge can be used by record labels, booking agencies or management companies to decide whether to focus on specific digital media outlets for marketing and promotion campaigns.

Identifying when and why traffic is increasing for a specific artist makes it possible to develop new revenue streams. Interviews with the different companies in this study shows that actors in the music industry have not yet started to use the full potential of increased online usage to sell ads or in other ways create new revenue streams. The first step for building these new business models is to understand the statistics and data and the presented system simplifies the process. By analyzing artist statistics it is possible for the music owner to understand why a certain artist might be more popular at MySpace than YouTube and this knowledge can then be used for focused promotion campaigns.

As revenues are moving from traditional media and tangible goods to music listening at digital services, we believe that the MIMS system will get more and more important as a tool for different actors within the music industry system. The handling of large amounts of data from digital services is especially important for collecting societies, who at the moment are struggling to find new ways for collecting and dividing revenues from public performance and remuneration online. The exploitation potential for new revenues becomes obvious when the firm can identify when, where and why fans are listening to their music, and makes it possible for the music industry to commercialize the traffic by placing ads or driving traffic to payment solutions for music.

References

- [1] Johansson D. (2006) Digital Gatekeepers Related to the Music Industry System
- [2] The NPD Group (2009) Continued CD Sales Declines in 2008, but Music Listening and Digital Downloads Increase
- [3] Edström-Frejman A. & Johansson D. (2008) Emerging and Conflicting Business Models for Music Content in the Digital Environment
- [4] Théberge P. (1999) Key Terms In Popular Music and Culture, p. 209 – 224.
- [5] Wikström P. (2006) Reluctantly Virtual – Modelling Copyright Industry Dynamics
- [6] Kusek D. & Leonhard G. (2005) The Future of Music – Manifesto for the
- [7] Fischer W. (2004) Promises to Keep

- [8] Bockstedt J. et al (2005) The Move to Artist-Led Online Music Distribution: Explaining Structural Changes in the Digital Music Market
- [9] Arvidsson K. (2007) Skivbolag i Sverige – musikföretagandets 100-åriga institutionalisering (Record companies in Sweden – a 100-year institutionalizing of the music company)
- [10] Eriksson L-E et al (2005) Broadband technologies transforming business models and challenging regulatory frameworks – lessons from the music industry
- [11] Styvén M (2007) Exploring the Online Music Market - Consumer Characteristics and Value Perceptions