Making Metadata: The Case of MusicBrainz

Jess Hemerly
jhemerly@ischool.berkeley.edu

May 5, 2011

Master of Information Management and Systems: Final Project

School of Information

University of California, Berkeley

Berkeley, CA 94720

Making Metadata: The Case of MusicBrainz

Jess Hemerly
School of Information
University of California, Berkeley
Berkeley, CA 94720
jhemerly@ischool.berkeley.edu

Index of Tables and Figures

Making Metadata: The Case of MusicBrainz

Jess Hemerly
School of Information
University of California, Berkeley
Berkeley, CA 94720
jhemerly@ischool.berkeley.edu

Summary

MusicBrainz is a "community music metadatabase" and an "open music encyclopedia" to which users contribute information about artists, releases, tracks, and other aspects of music toward the goal of creating a "comprehensive music site" [1,2]. As a peer-produced music metadatabase, MusicBrainz is a constructed cultural commons where users develop and distribute musical knowledge through the community website, an institution that supports and manages the pooling of metadata into a database [56].

Studying how MusicBrainz works and why people contribute builds on a growing body of research that seeks to understand how contributors cooperate to create and sustain constructed cultural commons. Understanding how these projects develop, thrive, and sometimes fail sheds light on potential solutions to collective action problems and other social dilemmas relating to cooperation, specifically in regard to information pools and collective knowledge systems. This case study provides a point of comparison between other constructed cultural commons as well as links between cultural taste, social habits, and peer-production. It also emphasizes the socio-cultural importance of music metadata.

This study employed both quantitative and qualitative research methods, beginning with a survey administered to the MusicBrainz community and data scraped from user profiles, followed by observation and qualitative interviews with registered users, called editors. Qualitative and quantitative data were analyzed and interpreted at the same time. The study sought to answer the following research questions about MusicBrainz:

- Contribution: Why do people contribute? Who are the MusicBrainz contributors? What characterizes editors' participation? How is contribution linked to cultural preference? How can one compare contributors' motivations in MusicBrainz to other constructed cultural commons?

- Music Information: How does MusicBrainz negotiate guidelines and standards? What is the relationship of MusicBrainz to other music resources? What is the role of metadata in music technology?

Findings are split into four categories: Demographics, including information about age, gender,

region, and overall editing statistics; How MusicBrainz Works, an overview of the technical components and governance structure; Patterns & Processes, which describes the links between musical taste and contribution, how MusicBrainz serves as a tool for discovery, and the ways editors' decisions mimic those made by information professionals; and Attitude & Motivation, which examines intrinsic and extrinsic motivations that drive users to contribute, from their belief in the philosophy of open source to a compulsion for accuracy and consistency.

I.   Introduction

MusicBrainz is a "community music metadatabase" in which users contribute information about artists, releases, tracks, and other data about music toward the goal of creating a "comprehensive music site" [1]. As a peer-produced music metadatabase, MusicBrainz is a constructed cultural commons where users develop and distribute musical knowledge through the community website, an institution that supports and manages the pooling of music metadata [56]. Examples of other well-known constructed cultural commons include Wikipedia, free and open source software (F/OSS) projects, patent pools, and jamband fan communities.

The factual user-contributed information about music in the database is public domain, while non-factual information (e.g., folksonomic tags and annotations) and the live data feed are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 license. In addition to contributing data, editors use mailing lists, forums, and Internet Relay Chat (IRC) to negotiate how the data should be structured or represented and to get assistance with norms and guidelines. The non-profit MetaBrainz foundation operates MusicBrainz, and receives money from donations and by licensing hourly updates of the live data feed to external music services, including Last.fm and BBC Music [49].

Users can access the data in MusicBrainz in several ways. First, a user can browse the data through the website by searching for specific artists, releases, and tracks or by browsing, following hyperlinks from one entity to another. Second, a user can download a piece of software known as a "tagger," which allows a user to scan music within his or her digital music collection and save the metadata tags to the files. Third, a user harnesses the Semantic Web capabilities of MusicBrainz by utilizing the XML web service to "query the database from any application that can parse XML" [3]. Finally, a user can set up his or her own MusicBrainz server to handle the live data feed of updates from the database.

Because the data is free to use, the only limitation on how a user may use it is that the factual data must remain in the public domain and the non-factual data and live data feed are governed by the Creative Commons A/NC/SA license—that is, the data cannot be repackaged and sold by another entity. The project's social contract promises that, "MusicBrainz will remain 100% free," and that the project's priorities are its "users, free content, and free software" [4]. This free public data makes MusicBrainz an information resource that functions as a public good [20].

This case study attempts to address two key sets of research questions, beginning with a set of descriptive questions about MusicBrainz: Who are the MusicBrainz contributors? What characterizes editors' participation? How is contribution linked to cultural preference? How can one compare contributors' motivations in MusicBrainz to other constructed cultural commons?

These descriptive questions intend to explore first key research question: Why and how do people contribute to MusicBrainz?

The second set of questions addresses the relationship of MusicBrainz to commons and peer-produced projects and to the ecology of music metadata and information about music available online: How does MusicBrainz negotiate guidelines and standards? What is the relationship of MusicBrainz to other music resources? What is the role of metadata in music technology?

This paper begins with the Background section, which first provides an overview of the challenges of metadata, specifically standardizing and organizing music metadata (A. The Problem of Music Metadata), and then briefly describes the history and structure of MusicBrainz (B. Why MusicBrainz?). The next section reviews relevant literature on cultural commons in relation to MusicBrainz (C. Collective Action and Constructed Cultural Commons). Section III, Methodology, describes the mixed methods research methodology employed in this case study. The paper then moves into the research findings in section IV, beginning with basic demographics and a detailed explanation of the technical infrastructure and community organization of MusicBrainz. Analysis and interpretation has been divided into two sections:

Patterns & Processes and Attitude & Motivation. Patterns & Processes includes findings relating to how editors use MusicBrainz to discover new music, how editors have created a new system that mimics work of information professionals, and how patterns of editor contribution relate to music acquisition and free time; links between Last.fm and MusicBrainz participation; and the efficacy of processes for consensus and resolution. Attitude & Motivation examines intrinsic and extrinsic motivations among MusicBrainz editors, including a widespread belief in the open source philosophy, the importance of reward and reputation, and enjoyment gained from the task of editing. The final subsection of the findings section discusses numerous areas for future work using the research findings.

II.   Background

A.   The Problem of Music Metadata

In 2008, Last.fm employee Richard Jones posted to the company's blog about a metadata problem he called "the Guns N' Roses Issue." Jones wrote, "Back in December I used Guns N' Roses to illustrate the music metadata problem by asking: Just how many ways to write "Guns N' Roses – Knockin' on Heaven's Door" are there?" [48]. He included a list of the top 100 ways that Last.fm users had represented the song in their music metadata, submitted to Last.fm through its "scrobbling" feature. Scrobbling software tracks, records, and sends information about music a user listens to on a computer or other device to Last.fm servers. At the time of Jones's post, Last.fm relied on accurate metadata to post the proper track and artist to a user's profile. The top 25 results from the list of 100 metadata variations are show in Figure 1: Top 25 Representations of "Knockin' On Heaven's Door" [48].

Figure 1: Top 25 Representations of "Knockin' On Heaven's Door" [48]

```
Guns N' Roses    -   Knockin' On Heaven's Door
Guns N' Roses    -   Knocking On Heavens Door
Guns 'N' Roses   -   Knockin' On Heaven's Door
Guns N' Roses    -   Knockin On Heavens Door
Guns N' Roses    -   Knockin' On Heavens Door
Guns N'roses     -   knockin on heavens door
Guns N' Roses    -   Knocking On Heaven's Door
Guns N Roses     -   Knockin' On Heaven's Door
Guns N Roses     -   Knockin On Heavens Door
Guns And Roses   -   Knocking On Heavens Door
Guns Nroses      -   Knockin On Heavens Door
Guns'n' Roses    -   Knockin' On Heaven's Door
Guns N Roses     -   Knocking On Heavens Door
Guns'n'Roses     -   Knockin' On Heaven's Door
Guns 'N Roses    -   Knockin' On Heaven's Door
Guns & Roses     -   Knockin' on Heaven's Door
Guns N'roses     -   Knockin' On Heaven's Door
Guns and Roses   -   Knockin' On Heaven's Door
Guns'n Roses     -   Knocking On Heavens Door
Guns 'n' Roses   -   Knockin' On Heavens Door
Aerosmith   -   Knocking On Heaven's Door
Guns 'n Roses    -   Knocking On Heaven's Door
Guns 'n' Roses   -   Knocking On Heavens Door
Guns N Roses     -   Knocking On Heaven's Door
Guns N' Roses    -   Knockin On Heaven's Door
```

Scrobbled data also builds a user's Last.fm library, and listening history data is the data upon which Last.fm generates music recommendations. The inconsistency in user data caused headaches for Last.fm, who spent a great deal of time working to resolve misspellings against correct representations.

In 2002, Professor Richard Smiraglia, then at the Palmer School of Library and Information Science at Long Island University, wrote about Music Information Retrieval (MIR), an area of

research and "very pragmatic implementation," and its implications for librarians and bibliographers. According to Smiraglia, "MIR embraces everything from 'query-by-humming' systems that allow a searcher to hum a tune for which the database returns an audio output, to the design of metadata structures and standard name-title-subject querying of bibliographic databases" [71]. This paper's case study focuses on design of metadata structures in a particular metadatabase, MusicBrainz.

Metadata has been defined in a number of ways by various institutions and organizations in the information organization space, from libraries to profitable companies. At the most general level, metadata is "data about data," but this basic definition is often supplemented with more nuanced descriptions. For example:

— Dublin Core Metadata Initiative (DCMI): "Metadata articulates a context for objects of interest—'resources' such as MP3 files, library books, or satellite images—in the form of 'resource descriptions'" [24].
— International Federation of Library Associations and Institutions (IFLA): "[Metadata] refers to any data used to aid the identification, description and location of networked electronic resources. Many different metadata formats exist, some quite simple in their description, others quite complex and rich" [46].
— National Information Standards Organization (NISO): "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource"[62:1].

Despite standardization attempts, such as Dublin Core Metadata Initiative, there exists no single authoritative format for metadata in the world at large, and even less authority control when it comes to multiple media types. Dr. Sherry Vellucci, now the Dean of Libraries at the University of New Hampshire wrote in 2001:

Authority control works well in the local library catalog, where the four success factors (well-defined boundary, application of principles and standardized practices, reference to authoritative lists, and highly trained professionals) can operate in a generally unimpeded environment; and while expansion of authority control into the cataloging arena faces more complex problems, the success factors are still operable because authority control remains within the common milieu of the library catalog. No such common environment exists, however, for the broader metadata world. [76:547]

Organizations like Dublin Core and NISO have worked toward developing metadata standards, but their application depends on a variety of uses [59]. And when it comes to multimedia for digital content, due the nature of free information exchange, even if standards did exist, the average user would only be exposed to those metadata fields common across multimedia players. And as the Guns 'N Roses issue illustrates, users do not always get those basic fields right. The digital metadata environment is the Wild West of information standards.

Music metadata is particularly tricky because it takes a number of forms depending on context: call numbers in a karaoke songbook; album and track listings with call numbers in a jukebox; and bibliographic information for where to find an album in a library. There are also multiple

instantiations of a single work, from cover songs to performances of classical work. With the shift to digital music, whether purchased through online retailers and subscription services, ripped from CDs, or acquired through file-sharing and BitTorrent communities, metadata has become more than just "data about data." In a digital context, metadata is *the* information that allows a user to locate a track or album within his or her music collection. Without metadata, a music fan is left with folders and files that cannot be identified within a music player— "artist01," "track 3," and "unknown album" tell a user nothing about the music contained within the listed file.

However, conceptually and socially, there is more to metadata than its role in MIR. Simply put, metadata is what we talk about when we talk about music. Smiraglia writes, "The musical work is received and interpreted within conventional cultural behaviors; it is the entire context of the musical work that has symbolic meaning" [71:755]. Metadata can include information beyond the artist, track, and album name, like an artist's repeated catch phrase, song lyrics, and even an artist's age at death. All of this information provides important contextual cues regarding the cultural and social identity of the music and its fans. Getting metadata wrong can have consequences ranging from social embarrassment—misidentifying an artist—to, as controversial royalty collection organization SoundExchange claims, artists not getting paid for Internet radio play [70]. Because of the infinite nature of metadata, when it comes to designing metadata structures and developing standards, choices must be made about what to include, what to exclude, and how the data should be structured and represented.

Form rules, the tools of authority or vocabulary control in bibliographic language, "ensure the consistency in description that is required for systematic access to bibliographic information" [73:54]. According to Svenonius, there are two types of terms that constitute a controlled vocabulary: terms that are "derived" and terms that are "assigned." Derived terms are "descriptive metadata elements," taken "as-is from documents"; the assigned terms are "organizing metadata," providing bibliographic structure and applying a controlled vocabulary that "facilitates the establishing of relationships among entities"[73:56]. In the case of music metadata for music, the terms are usually a mix of derived—track number, artist—and assigned—genre, unique identifier, or call number if in a library—terms.

The National Information Standards Organization (NISO) breaks metadata down into finer categories of structural, descriptive, and administrative metadata:

> *Descriptive metadata* describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
>
> *Structural metadata* indicates how compound objects are put together, for example, how pages are ordered to form chapters.
>
> *Administrative metadata* provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. [62:1]

But these are not always mutually exclusive when it comes to music. For example, the record label on which an album is released can serve as both administrative and descriptive metadata. In the case of small labels like Daptone Records, which produces soul artists, and Matador, which housed some of the 1990s major independent rock acts like Yo La Tengo and Pavement, the administrative detail of label association provides a listener clues as to the nature of the music itself, thus being both administrative and descriptive.

Today, there exist two major commercial sources for music metadata: Gracenote and Rovi's AllMusic. (A third company, Echo Nest, is quickly becoming a player as well, but has not yet established the same level of corporate relationships as the other two.) AllMusic began as AllMusicGuide, a reference book that listed music information, all of which was manually entered from CDs. The CDs were (and still are) stored in a massive warehouse in Michigan [67]. As the project grew, they created a website to show potential commercial customers the extent of the data they possessed. AllMusic employed editors to manually add data, to classify music by genre and style, to relate similar artists to one another, and to produce editorial content like biographies and reviews. Today, AllMusic provides metadata to several notable commercial digital music services like iTunes, Napster, and Borders. Until very recently, AllMusic's data was inaccessible to those who wanted to use the raw data but were not commercial clients. However, AllMusic recently launched an API service that allows developers to play around with and use the extensive data that AllMusic has collected over time [67].

Gracenote also has an API, with different licenses for commercial and non-commercial developers. However, their SDK source code is licensed under the "Gracenote Open Source License," which is "nonviral," unlike copyleft licenses like the GNU General Public License [42]. iTunes users who insert a CD into their computers and use iTunes' built-in CD identification are familiar with Gracenote, as iTunes service reads the CD's table of contents and matches it to table of contents in Gracenote's Compact Disc Database (CDDB). iTunes users can submit metadata they have manually added to an unrecognized CD, but it is not added to CDDB immediately. Gracenote also offers music analysis services, which scan an audio track and match the waveform to a different Gracenote database. TuneUp, a San Francisco-based start-up, provides a Gracenote-powered "cleaning" service for users who wish to use the database to assign the proper metadata to their entire (iTunes or Windows Media Player) music collection, including album art. However, while iTunes CDDB access is free, TuneUp is only free up to 50 "cleans" [75]. Beyond that, users must pay to access the technology. And CDDB also has its limitations. Because CDDB relies on table of contents lookup, someone who has inserted, say, a mix CD, will not receive metadata.

Any data corrections must be submitted to these companies, and filtered through editors. Submissions of new releases at AllMusic must also go through editorial staff. Because of the nature of the music market and the manpower required to enter the data, lesser-known releases may be de-prioritized in favor of releases in demand by commercial clients [67]. Neither of the two major proprietary sources is peer-produced or peer-edited, and neither source is truly free.

B.   Why MusicBrainz?

Project founder Robert Kaye established MusicBrainz after Gracenote purchased and privatized the CDDB, which was once a peer-produced information resource, or an information pool. As Kaye puts it, "I typed in probably about 200 of my own CDs and magically, when [Gracenote] took to private, I didn't get a check for my efforts" [49]. According to Kaye, a friend encouraged him to start his own open project and in 1998, the CD Index was born. Thanks to a post on Slashdot, the site became hugely popular overnight, with 3,000 users and 10,000 entries in the database. But the community was not quite what Kaye imagined, and, disappointed with the way users treated each other, Kaye began work on a new version of the database. "I knew that CDs were obviously not the future, and MP3 or digital audio was going to be representing the future," says Kaye. "So I am going to rebrand this as MB and turn it into a music encyclopedia over the next few years" [49]. As a reaction to Gracenote's privatization of the once public CDDB, Kaye vowed to ensure that MusicBrainz would be an open source project, selecting a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 license for the non-factual data and the live data feed.

MusicBrainz is not the only open music metadata resource. Among MusicBrainz users, the most notable are Discogs.com and FreeDB. All three are peer-produced resources, and MusicBrainz provides importing capability from the FreeDB database. However, MusicBrainz editors in this study questioned the data quality of Discogs and FreeDB, but are especially wary of FreeDB. The MusicBrainz community has driven discussion and negotiation of style guidelines and data representation and do not see the same community dedication to quality in other sources. The MusicBrainz community has been instrumental in the design of the Next Generation Schema, currently in the testing phase, which aims to better represent music metadata for all genres. The community is the authority, controlling the metadata standards.

The database started with only the basic metadata information like that found in CDDB: album, track, and artist names. But in 2005, MusicBrainz added an Advanced Relationships feature that dramatically increased the amount of information that could be included in database. This addition of links between artists, performers, producers, and outbound links on the web positioned MusicBrainz to be more like a music-specific encyclopedia of structural metadata, and pushes on traditional notions of music metadata to include many more fields and aspects than those basic pieces of information necessary for finding a file in a music collection.

Thus, the metadata in MusicBrainz includes the different kinds of metadata described by NISO: descriptive metadata such as name and title; structural metadata, in the advanced relationships; and administrative metadata, like PUIDs (acoustic fingerprints) and MusicBrainz IDs, MusicBrainz-specific unique identifiers. The Million Song Dataset, recently released by EchoNest, includes these MusicBrainz IDs, making MusicBrainz part of a rich dataset for researchers and technology designers. EchoNest & LabROSA explain, "Fields 'year', 'artist_mbtags' and 'artist_mbtags_count' have been extracted from the MusicBrainz. We used a local copy of the server, our version is this branch, the data dumps were of December 4th, 2010. Note that the field 'artist_mbid' is provided by the Echo Nest API" [25].

MusicBrainz data is highly structured, and, as noted in the introduction, provides an XML web service and a soon-to-be-revived RDF format so that technologies can more easily access, use,

and remix the data. Because of the high value in encoding standards for metadata, like XML, and structural standards, like RDF [76:553], MusicBrainz is positioned to be a primary source for music information as Semantic Web technologies develop, enabling researchers and designers to experiment with machines, systems, and agents that combine and leverage the information contributed to the collective knowledge system [43:446]. Catherine Marshall and Frank Shipman note that, in terms of the Semantic Web as metadata, "Metadata is not simply a description of the information contained in a work or web page; the choice of metadata scheme also signifies community membership" [59:5]. As a metadata resource for the Semantic Web, a designer's decision to use MusicBrainz could indicate an alignment with the principles and objectives of the project, and, generally, with Creative Commons and open source philosophies.

Marshall and Shipman also discuss the tradeoffs involved when using amateurs to reduce metadata costs and the importance of trust in the authority of a metadata source. With MusicBrainz, the negotiation of style guidelines and the metadata's entry into the database are all tasks performed by a community of non-professionals. However, these non-professionals are dedicated to high quality, and, as this paper will explore in great detail, have developed standards and authority control to achieve that level of quality. The data is of high enough quality that, in 2007, Last.fm and BBC Music reached commercial licensing agreements with MusicBrainz, and the project has provided metadata to both sites since.

Researchers studying personal music collection habits found that such "extra-musical" documents, including lyrics, provide additional background and augment the "listening experience" [23:6]. Some users desire richer metadata than others, ranging from track length and album art to lyrics. When it comes to metadata, the same researchers concluded, "as rich a set as possible should be available, with the user able to select the fields of interest for display"[23:6]. This presents a unique opportunity to link musical fandom and peer production, as well as for MusicBrainz to serve as a discovery tool.

As mentioned above, the scope of the data in MusicBrainz goes beyond basic track listings. Knees et. al. note that traditional music search engines are limited "to a rather small set of meta-data, whereas the musical, or more general, the cultural context of music pieces if not captured" [51:447]. Advanced relationships within MusicBrainz are one way that the project seeks to record contextual cultural context of music by including lists of performers in a band to notable producers and recording eras. When designing an advanced music search engine, Knees et. al. used the "underlying semantic relationships" in MusicBrainz to enhance music search, recognizing that these relationships result from information that constructs a more "extensive set of manually annotated metadata" [51:448]. MusicBrainz has harnessed the social power of music and metadata to build a comprehensive database of administrative, descriptive, structural, and contextual metadata that grows every day.

MusicBrainz has a field for editors to add descriptive tags about music, but MusicBrainz does not include official genre or style information. Genres are highly dependent on cultural context and are simply too subjective for the community to agree on. For example, "world music" has been a consistently controversial genre label because it is Western-centric, a catch-all for things that are non-Western or "ethnic." [68:9]. Additionally, as a quick glimpse at AllMusic's carefully crafted styles [5] or SoundUnwound's user-generated tags [6] show, genres range from the very

broad to the very narrow and can describe an era, a mood, a type of rhythm, a region of origin, and more. MusicBrainz has a tag field for such information, but does not impose any official genre on artists.

C.   Collective Action and Constructed Cultural Commons

Garett Hardin's *Tragedy of the Commons* relies upon the example of a pasture where, after years of natural forces limiting the amount of animals on the pasture, "social stability" is achieved and herdsman will seek to maximize their gain. According to Hardin, the instinctual action would be for each herdsman to continue to add an animal to his herd until the tragedy of the commons is reached, where the pasture is so overpopulated that it can no longer sustain the herds. Hardin's tragedy of the commons "involves resource users over-exploiting a resource and imposing mutual externalities upon *each other*" [45:3], and his argument focuses on the "population problem," a problem he considered to be one of the "no technical solution" class.

But within Hardin's tragedy, Elinor Ostrom found the potential for success. Ostrom developed models for investigating natural resource commons, and applied them number of case studies. Her studies showed that, contrary to Hardin's pessimistic assumptions, it is possible for actors in the commons to overcome collective action problems [64]—specifically social dilemma games whose rational outcomes dissuade cooperation, like the Prisoner's Dilemma [15]—and establish successful and durable commons-pooled resources.

> Collective-action problems occur when individuals choose actions—such as whether to build and maintain an irrigation system—in an interdependent situation. If each individual in such situations selects strategies based on a calculus that maximizes short-term benefits to self, individuals will take actions that generate lower joint outcomes than could have been achieved. [65:155]

According to Ostrom, Nash equilibrium (where "actors interacting with each one another each choose their best strategy given the strategies that all other actors have chosen" [57:368]) never yields the "socially optimal outcome" for a single iteration of the game. Instead, the socially optimal outcome can only be achieved when players choose to cooperate by "selecting strategies other than those prescribed by Nash equilibrium" [65:155]. Despite the outcome prescribed by Nash equilibrium, Ostrom has shown that commons can exist because people are often willing to forego the "best strategy" and cooperate to share resources.

Commons have evolved from the natural resource pools like the Maine lobster fishery [56]. Information resources can be collected in an information pool [20], enabled by several key features of digital technologies. According to the Bimber et. al., technologies can help to overcome collective action problems by enabling the three basic tasks required for successful collective action. They write, "The need to accumulate resources in order to bear the costs of acquiring information about interests, the costs of distributing messages, and the labor and material costs of coordination are diminished substantially under certain circumstances by the availability of new technologies"[17:374]. Information resources are also unique in that information is neither a purely public nor a purely private good: "It is a good that people simply

do not use up, as they do other goods" [53:9].

Madison et. al. narrow the category of information pools to what they call "constructed cultural commons"[56]. Constructed cultural commons are "environments for developing and distributing cultural and scientific knowledge through institutions that support pooling and sharing that knowledge in a managed way, much as a natural resource commons refers to the type of managed sharing environment for natural resources" [56:659]. The authors also note that, "unlike resources in the natural world, resources of information and expression must be created before they can be shared" [56:672]. So the constructed cultural commons relies on cooperation both to create and to share the resource through the process of peer-production. Examples of constructed cultural commons in Madison et. al.'s paper include patent pools, open source software, Wikipedia, the Associated Press, and jamband communities, where members produce and share recordings of live performances.

MusicBrainz is a constructed cultural commons for music metadata. The project's infrastructure is funded by donations and clients who subscribe to "timely and convenient data access" for commercial use [7]; the database has constructed one release at a time from the contribution of users; limited intellectual property constraints control who can use the data and how; and the community has developed guidelines for what information can be added and how. Here, licenses are the only intellectual property constraints: the Creative Commons (Non-Commercial-Share-Alike 2.0) license governs non-factual information and the live data feed, to prevent privatization and exploitation of the database; and the General Public License 2.0 or later for Picard, the tagging software that allow users to easily add MusicBrainz metadata to their digital music files.

The MusicBrainz introduction page reads, "Any user that contributes to MusicBrainz should be aware that their contributions will be made available to the public under the licenses described below" [3]. While it's not a public good in the purest definition of the phrase, the data MusicBrainz makes available to the pubic is a public good. Because the public is free to use and take the information, MusicBrainz faces a common collective action problem: free-riding. As Cheshire and Antin write, "As is the case in public good problems, individuals who participate in information pools must overcome the temptation to free-ride (consuming the public good without making a contribution to it)" [20:706].

Because all users can access the same data at the same time, an information resource that is a cultural commons appears nonrivalrous. However, nonrivalry is limited even in the digital constructed cultural commons. The resources—in this case editors—who contribute to the information resource are exhaustible, and "the possibility of physical co-use does not eliminate the possibility of an originating donor monetarily, emotionally, or practically losing the ability to use what he or she helped to create" [41:739]. And while free-riders will not deplete the resource, the commons still suffers if all users free-ride [17]. Free-riding would only cease to be a problem if an information resource was fully complete, where no more information could be added, thus the commons could no longer grow. But even if all fields were completed for existing entities, new music is constantly being created and released, so a database whose mission it is to capture data about all music can never be truly complete. As a result, MusicBrainz is susceptible to free-riding.

MusicBrainz licenses prevent free-riding by commercial entities, which allays fears about an outside entity coming in and privatizing the database [41:686], as Gracenote did with CDDB. But non-commercial, individual users are still able to free-ride. If a user tagging his or her music collection only tags files that are already in MusicBrainz and does not add any new data, the cultural commons suffers because the user took from the database what was there but withheld information he/she had that the commons did not. Thus, there must exist motivations that drive people to cooperate—in this case to edit or add to the information in the database.

Madison et. al. also note that, "Cultural commons are also nested within and interact with more complex systems of natural and socially constructed environments" [56:681]. MusicBrainz is a part of two main larger systems: the ecosystem of semantically enabled, peer-produced information online and the expansive realm of the music industry [74]. MusicBrainz is an open source alternative to proprietary music data sources, and is a major peer-produced music-specific resource on the web. Its structure, its governance, and its existence depend on its position in these larger socially constructed environments.

What is it that motivates people to construct this cultural commons, to cooperate and contribute to the music information resource? Studies have been conducted on a various aspects of commons-based, peer-production communities, notably free and open source software projects [40,44,55,63] and Wikipedia [14,52,66,69,77]. However, no major studies have looked at the contributions of individuals in a community working to build a metadatabase. Where Wikipedians play the role of encyclopaedists and open source software contributors fill in for highly structure commercial organizations, MusicBrainz contributors play the role of information scientists as digital librarians, standards-setters, and catalogers of music.

In *Free and Open Source Software,* Lakhani and Wolf present the analysis of a quantitative survey on motivations and contribution in open source software projects in a paper entitled "Why Hackers Do What They Do? Understanding Motivation and Effort in Free and Open Source Software Projects" [55]. The authors categorize motivation into intrinsic and extrinsic, with different sub-types of motivations within each category. Within the intrinsic motivation dimension there are obligation/community-based motivations and enjoyment-based motivations. Within the extrinsic motivation dimension there are immediate payoffs in the form of satisfying a user need and receiving monetary compensation for work, and delayed benefits that include career advancement and building professional skills. This motivational framework provided an appropriate model for constructing survey questions and interpreting findings.

One of the immediate payoffs is particularly relevant in the case of MusicBrainz: "user needs to solve a particular software problem may also drive participation in F/OSS projects"[55:7]. Delayed benefits were not a focus of this research since none of the editors in the study discussed any direct link between their careers and MusicBrainz—that is, while several of the subjects are computer programmers and engineers, none are metadata specialists or information scientists. However, interviews did reveal that MetaBrainz hired some active contributors because of their contributions to the project, turning it into a career (this is discussed further in Section IV). There are only a total of 3 paid full-time employees and one part-time employee, so pay is not a motivation for most users.

In terms of Lakhani and Wolf's intrinsic motivations, enjoyment-based motivation is important in the case of MusicBrainz, but potentially more important is the latter motivation, where "individuals may be socialized into acting appropriately and in a manner consistent with the norms of a group. Thus the goal to act consistently within the norms of a group can trigger a normative frame of action" [55:5].

As the Patterns & Processes and Attitude & Motivation sections of this paper will show, the motivations Lakhani and Wolf identified are highly relevant in the case of MusicBrainz. However, these motivations are complicated and complemented by the matter of musical taste. Musical taste is often the gateway to participation, and human connection with music is highly social and highly emotional [58,61]. Thus, a social and emotional connection to music drives an information need and shapes editors' contributions. In addition to the motivations Lakhani and Wolf identify, an emotional connection with music serves as an intrinsic motivation for participation in this metadata commons, along with a self-described compulsive need for the data to be correct.

III. Methodology

The purpose of this two-phase, sequential mixed methods study was to gather quantitative data from a sample of MusicBrainz users to get a descriptive overview of how and why they use MusicBrainz, and then to follow-up with qualitative interviews in order to more deeply explore the motivations and patterns of participation among active editors in the community. Communication channels used by the community were also examined, including forum posts and chat logs. Throughout the study, I contributed to the database as an editor—a participant-observer—in order to stay abreast of any technical or representational changes to the database and to experience the commons as a new contributor. The University of California, Berkeley's Center for Protection of Human Subjects approved this study under protocol number 2010-03-937.

The qualitative research is given priority, but the diverse data types allow me to address different problems and questions, especially given the pragmatic nature of the research problem [22]. The quantitative research has addressed questions regarding general demographics, overall attitude, and the relationship of MusicBrainz to other sources. It also shows that certain claims made in interviews are applicable to the broader survey sample. The qualitative research facilitated a deep exploration of processes and governance mechanisms, habits of contribution, motivations, and musical taste. The survey and scraped profile data provided a data set with which to find interesting relationships and possible dependencies between different factors of contribution.

However, there are limitations to the meaning that surveys can uncover and clarify, even with the inclusion of open-ended questions [72]. While the survey is the instrument for the quantitative aspect of the research, the survey's design was limited by the questions created at the time of construction. Conversation-driven qualitative interviews allow for revision of interview guide questions before and after each interview in order to expand the breadth of issues—even including notes about some of the tangential themes—that should be covered in future interviews.

Integration of the two data types occurred at the data analysis stage, but quantitative data informed the qualitative interview guide, and qualitative data informed quantitative analysis. This study is specific to this particular community of users and cannot be generalized to all constructed cultural commons. However, the results do inform our understanding of how people in other user-generated content communities may think and feel in regard to their contributions, and, as a focused case study, adds to existing work in this area of inquiry. The great detail to which this study has gone provides a solid piece of work for comparison with similar case studies of other cultural commons.

A.   Quantitative Methods

The primary research questions for the quantitative research are: What motivates editors to contribute to MusicBrainz? What are the shared characteristics of the contributors? How does cultural production of MusicBrainz relate to musical fandom?

Survey Design and Implementation

The questionnaire administered to the MusicBrainz community was designed under the guidance of Professor Coye Cheshire. In order to attract respondents, MusicBrainz founder Robert Kaye provided input and later posted the survey link to the official MusicBrainz blog, a source of news and updates for the community, in a post titled "Please take the MusicBrainz Survey!" [50]. A link to the survey was also posted in the MusicBrainz forums, where users of all activity levels go for guidance on edits and discussion about community projects, and the MusicBrainz style and user mailing lists, active communication channels for MusicBrainz contributors. A link to the survey appeared on the study's observational WordPress blog as well.

The sample was not truly random; it was a purposive sample consisting of people who followed the community's various communication channels [16:176]. As such, it is biased toward editors more active in the community than those who may simply use MusicBrainz through one of the project's supported metadata taggers.

The survey consisted of two main sections: questions for all users and questions for users who have registered accounts, known in the community and henceforth identified in this report as "editors." All respondents had to confirm that they were at least 18 years of age in order to access the survey questions. The non-registered user section, MusicBrainz Usage, included questions about how respondents found and use MusicBrainz, whether or not they have discovered artists through MusicBrainz, and basic questions about habits for finding information about music online. Questions for editors aimed to understand the tasks the users perform when contributing and, using Likert-type questions, general attitudes and behaviors regarding other editors, resources, community practices, and processes. A section of questions targeted at registered editors requested information about username, registration date, and editor status. Finally, an optional open-ended field allowed editors to volunteer their email addresses if they were willing to participate in a follow-up qualitative interview. All questions were optional with the exception of the age requirement question at the beginning. All respondents above 18, registered editors or otherwise, were asked to answer basic demographic questions about their age, region, and gender (see Appendix for the full survey).

The survey was administered through open source survey software LimeSurvey hosted on UC Berkeley's School of Information (I School) web space. Several I School students familiar with MusicBrainz and one very active MusicBrainz user tested the pilot survey, and questions were reconfigured based on their input. The survey was open to responses from anyone who came found the link from July 8, 2010 until 12:00am, July 24, 2010.

A total of 266 respondents over the age of 18 responded to the survey, with nine respondents who did not complete the survey in its entirety. These respondents were removed from the dataset for analysis, along with responses from the project's founder, leaving a total of 256 responses. Of this number, 241 respondents are registered MusicBrainz editors. This distinction is important, as anyone can be a user of the data, but only those who have registered usernames can contribute to the data. To put the sample into perspective, during the same time period, 2237 editors made at least one edit in the database, and 557 editors made 10 or more edits (data provided by MusicBrainz).

114 editors voluntarily provided an email address for a follow-up interview, and 186 of the 241 registered editors provided valid usernames. Figure 2 shows responses of those who provided usernames to the categorical question "When did you last log in to MusicBrainz?" Of the editors who provided a username and answered this question (N = 184), 111 editors logged in to MusicBrainz the same day they completed the survey.

Figure 2: When did you last log in to MusicBrainz?
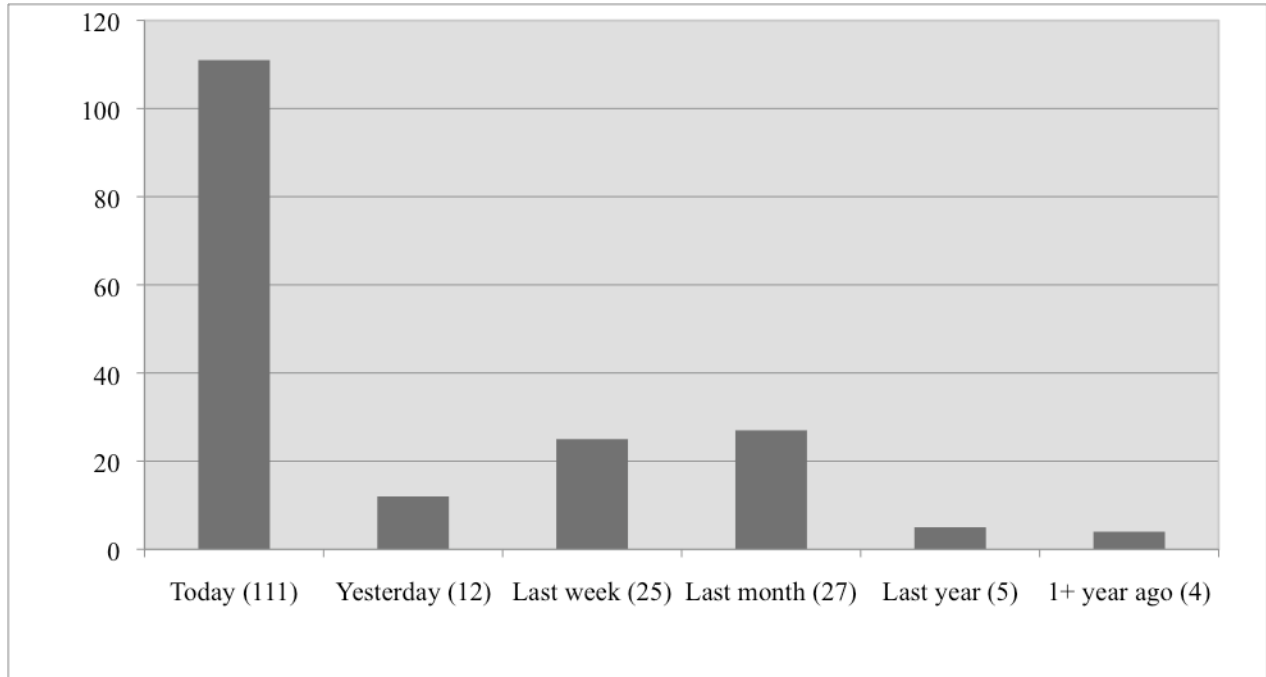
All quantitative data was analyzed and tabulated using Stata 11.

Python Scraping & Data Gathering

Given the large number of usernames collected, two Python scripts were created to gather more information about registered editors. They are discussed in the following sections.

Profile Scraper

User profiles on MusicBrainz display statistics for editors' contributions to the site over time. MusicBrainz editors can also "subscribe" to artists and other editors in order to receive email alerts when changes are made to a particular artist of interest. MusicBrainz displays the number of subscriptions on an editor's profile page. Using regular expressions, the profile scraping Python script cycled through the list of usernames, connected to each editor's profile page via HTTP request, extracted the edit, vote, and subscription statistics, and saved the information to a .CSV file. These numbers were then copied to the master survey spreadsheet to correspond with the usernames. This data collection occurred on March 28, 2011, and an error check of lines in the .CSV against editor profiles ensured that the correct information had been scraped.

Two usernames were removed from the list, one because the account could not be found, and the other because the username included characters that couldn't be formatted properly to perform the HTTP request.

IP Address Lookup
In order to complete some of these interviews and meet the requirements of a graduate-level qualitative methods class, it was necessary to find a way to find editors the Bay Area to interview in person. Editors identified their continental regions on the survey, but not for specific city/state/province/country locations. However, since IP addresses were recorded in the survey to prevent duplicates, an IP lookup service could be used to find their location. Using a list consisting only of the North American editors, a Python script was created to looked up their IP addresses at *http://www.melissadata.com/lookups/iplocation.asp?ipaddress* and return the location (city, state/province, country). Locations were then copied back into the .CSV file, creating a quick-glance record of North American editors' locations.

B.   Qualitative Methods

The primary research questions for the purposes of the qualitative work are: How is the musical taste of editors reflected in the contributions make to the site? What kinds of processes are in place for negotiation and decision-making within the community? What makes MusicBrainz a project worth editors' time? How do editors contribute? How are community processes structured and executed?

Participant Observation

Exploration of MusicBrainz began from the perspective a contributing editor, making small contributions to the database to learn how it worked and to get a sense of what kind of community existed, how they communicated, and what tasks editors perform. A blog of observations was kept early on, including notes about edits, interactions with other editors, and screenshots about questions. A few editors found the blog, a link to which was included in my MusicBrainz profile. Editors communicated with me as a new editor via the comments section on the blog and via edit notes on the MusicBrainz website.

Editor Interviews

A total of 24 editors who provided their email addresses for follow-up interviews were contacted, beginning with those editors whose IP addresses indicated a San Francisco Bay Area location because of the requirements for a qualitative methods class. All users who provided their email addresses were considered next, and at least one user in every geographic region indicated on the survey was contacted. Due to the small percentage of female respondents, attempts were made to arrange an interview with each of the editors who responded to the gender question as "Female," but none of the female editors responded to emails. However, one woman mentioned by several of my interview subjects was available for an interview and became a part of the study.

A total of 14 editor interviews were conducted over various communication channels depending on what was most convenient for the research subject (see Table 1: Completed Editor Interviews). Consent forms were collected from every respondent. The consent form provided two levels of consent: first, was a general consent to participate in the interview, and second,

consent to use the editor's username in any subsequent papers and presentations. All but one editor consented to the use of their username, but all editors were anonymized for inclusion in this paper. The same interview guide was utilized for each interview to ensure that the topic areas were covered. As qualitative methodology allows, questions were not asked precisely the same way each time; instead, conversation drove the order and way in which questions were asked [54].

Table 1: Completed Editor Interviews

Interviewee	Date	Type
Editor1	10/20/10	In person
Editor2	10/21/10	In person
Editor3	11/19/10	Skype (audio only)
Editor4	11/23/10	Skype (audio only)
Editor5	1/25/11	Skype (audio only)
Editor6	1/27/11	Skype (audio only)
Editor7	1/28/11	Skype (audio & video)
Editor8	1/31/11	Skype (audio only)
Editor9	1/21/11	Skype (audio only)
Editor10	2/8/11	Skype (audio only)
Editor11	2/17/11	Google Chat
Editor12	2/18/11	Jabber
Editor13	1/21/11	Skype (audio & messaging)
Editor14	3/8/11	Skype (messaging only)

Transcripts were coded following the interviews to identify major themes and patterns. Initial coding involved going line-by-line and inserting codes in half of the transcripts [19]. These codes were then translated to themes, each receiving a separate tab in an Excel document. The remaining half of the interviews were read line-by-line, with relevant quotes pasted into the Excel tabs corresponding to the appropriate themes. Once stories emerged from the themes, they were then combined as appropriate and formed the structure for this paper's Findings section. Supporting data was added to each section as appropriate, all culled from the interview transcripts. The coding uncovered patterns across interviews, and teased out some of the "possibly ambiguous and contradictory meanings" in subjects' answers [54:34].

Chat Log Analysis
Active MusicBrainz editors use a variety of information channels to communicate and coordinate. These channels include mailing lists, forums, and Internet Relay Chat (IRC), the most popular of which is *#musicbrainz* on freenode.net. MusicBrainz makes the chat logs available on the website, so in addition to using IRC to contact one editor, chat logs for the months of February were collected and used to get a general sense of how editors communicated with each other in that channel and what kinds of discussions occurred.

Expert Interviews
In order to understand how MusicBrainz compares to other music information resources and fits into the digital music landscape at large, several expert interviews with industry professionals were conducted. The first of these was Robert Kaye, the founder of MusicBrainz, followed by

Ian Rogers (CEO, Topspin Media); Brian Trenz (Senior Developer, Rovi); and Linda Quach (Public Relations, Rovi). Attempts were made to contact other experts but scheduling conflicts made it impossible.

IV. Findings

A. Demographic Overview

The gender makeup of the age groups of respondents is shown by percentage in Figure 3. Of the 248 respondents who provided their gender, seven were female and 241 male. (Of the users who answered "yes" to "Are you a registered user?", 227 editors are male and seven are female.) This indicates that my sample is predominately male, and editors suggested in interviews that there are very few women actively using the site. The age group with the most total respondents is 22–34 (155 male, 2 female), followed by 35–44 (50 male, 2 female). The 18–21 group (18 male, 1 female) and 45–54 group (15 male, 1 female) are close to each other for third and fourth, followed by 55-64 (2 male, 1 female) and 65+ (1 male) respectively.

Figure 3: Gender Makeup of Age Groups (All Respondents)
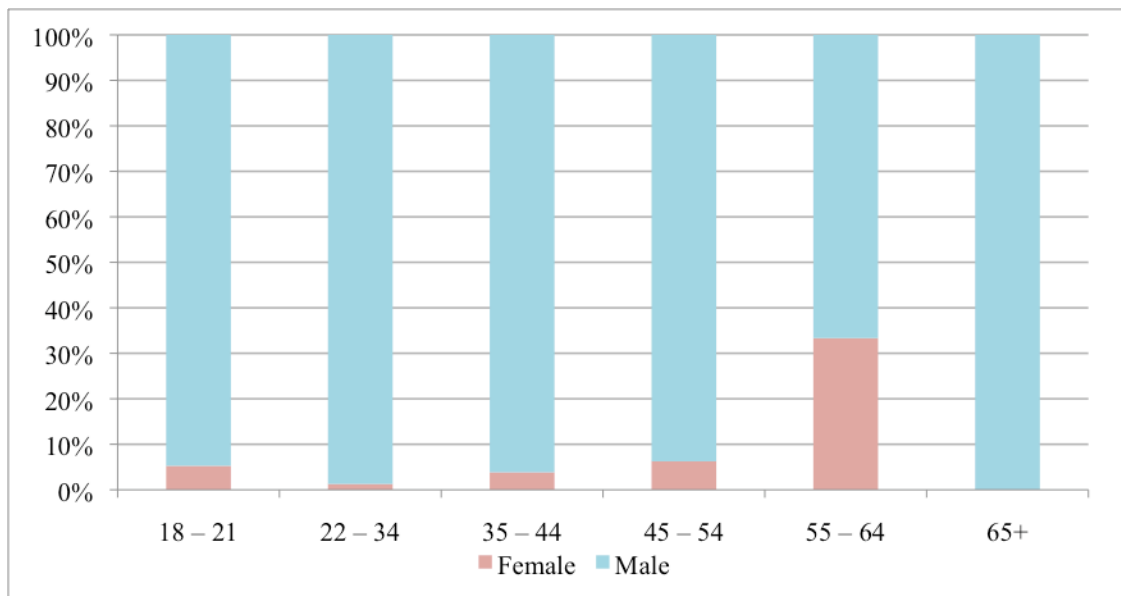
Each respondent was also asked to identify his or her region by selecting one choice from a mutually exclusive list of regions, and 257 provided a region. The percentage breakdown of regions is shown in Figure 4, with Europe having the largest share of the sample (153 respondents), followed by North America (78 respondents), Australia/New Zealand (12 respondents), South America (8 respondents), and Asia (5 respondents). Africa and Antarctica were also answer choices but none of the respondents selected those regions. Robert Kaye noted that an editor in Antarctica operates a MusicBrainz server but that editor did not respond to the survey.

Figure 4: Respondents by Region
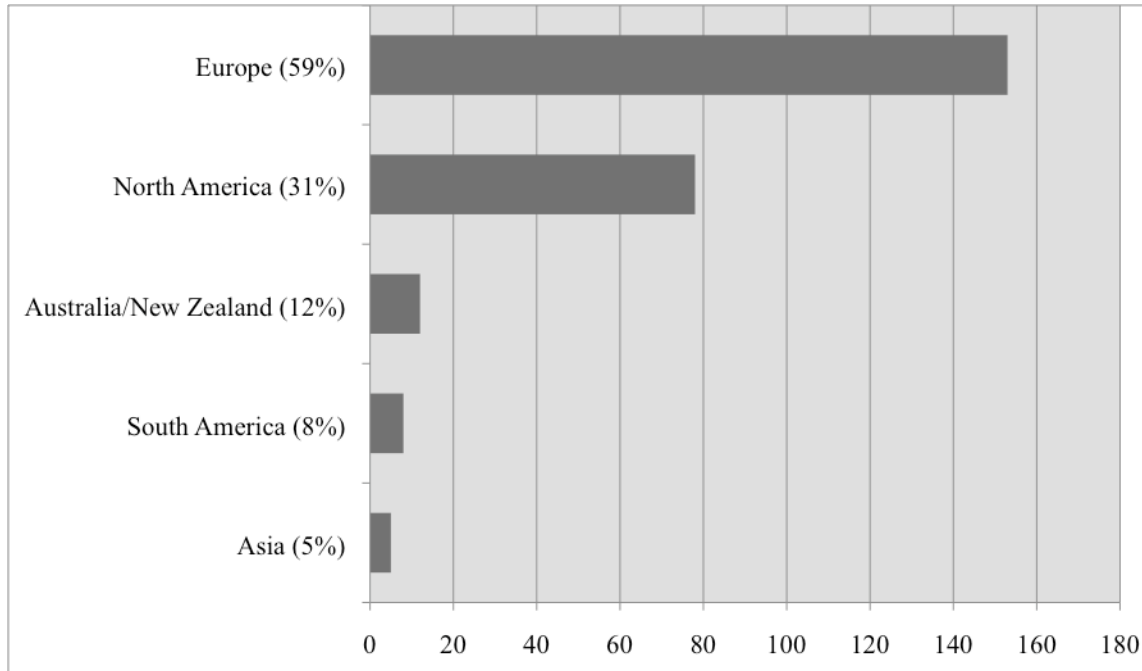
Given a list of 184 usernames, the Python script scraped statistics from user profiles as shown in Table 2: Summary of Scraped Profile Data. The information includes Artist Subscriptions ($\mu = 763.68$), Edits Entered ($\mu = 13,591.72$), Edits Accepted ($\mu = 3,864.179$), Edits Voted Down ($\mu = 16.1087$), and Edits Failed for Other Reasons ($\mu = 17.57609$). I did not remove outliers from these statistics as they are meant simply to illustrate the range of contribution within my sample. These numbers show that users who offered usernames rage from very few contributions to many.

Table 2: Summary of Scraped Profile Data

	N	Mean	SD	Min	Max	Skewness	Kurtosis
Artist Subscriptions	184	763.68	3,384.604	0	42613	10.57496	127.8038
Edits Entered	184	13,591.72	46,793.73	3	494,857	7.325237	67.4803
Edits Accepted	184	3,864.179	6,150.844	1	56,660	4.798738	35.95922
Edits Voted Down	184	16.1087	35.73321	0	370	6.268407	55.74809
Edits Failed for Other Reasons	184	17.57609	34.00586	0	264	4.505681	28.02504

A bar graph showing the total number of edits of users who offered usernames is shown in Figure 5. The ranges have been loosely based on percentiles in the edit data. The 25th percentile is 372; 50th percentile is 1,398 edits; the 75th percentile is 7,422 edits; 90th percentile is 27,297 edits; 95th percentile is 62,249 edits; and the 99th percentile is 257,422 edits. More users in this sample have individually made a fewer number of edits, while only a small number of editors have made over 20,000 edits.

Figure 5: Total Edits Entered (Respondents who provided usernames)
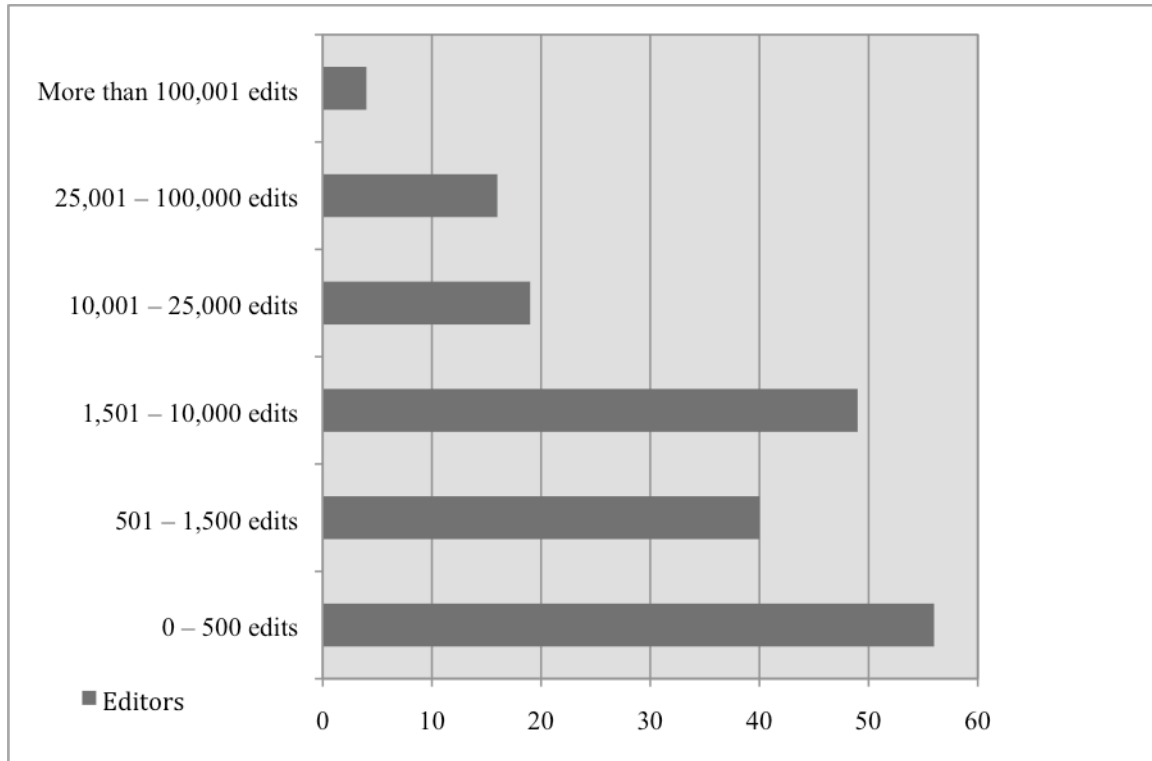
In order to work with the length of time that users in my sample have been members of the site, the variable *registerdate* was used. The date format in the .CSV file was changed from mm/dd/yy to Stata's preferred format, dd[mmm]yyyy, where [mmm] is the first three letters of the name of the month. For example, an entry of 01/25/06 became 25Jan06. Of the 255 respondents who said they are registered users, 209 of them provided their date of registration. Two respondents provided dates in the future. One of these respondents provided a username, so the correct data was manually copied into the data set from the editor's profile page. The other editor did not provide a username, so that editor's date was dropped from the variable. This leaves a total of 208 respondents with valid dates of registration.

Using Stata's date functions, number of years an editor has been a registered user was calculated from April 4, 2011 to generate the numeric variable *memlength* (summarized in Table 3, rounded to the ten-thousandths place). The mean number of years respondents have had registered accounts is 4.5622 years, and with a standard deviation of 2.0765, values are tightly clustered around the mean. The maximum number of years is 10.7871, and the minimum is 0.2683, or approximately 98 days.

Table 3: Years Registered

	N	Mean	SD	Min	Max	Skewness	Kurtosis
Years Registered	207	4.5622 yrs	2.0765	0.2683	10.7871	-0.07	2.6397

From the variable *memlength*, the variable *membyyr* was created to separate the respondents into

the number of years they have been members. Results shown in
Figure 6, with number of respondents on the y-axis and years registered on the x-axis.

Figure 6: Respondents by Number of Years Registered
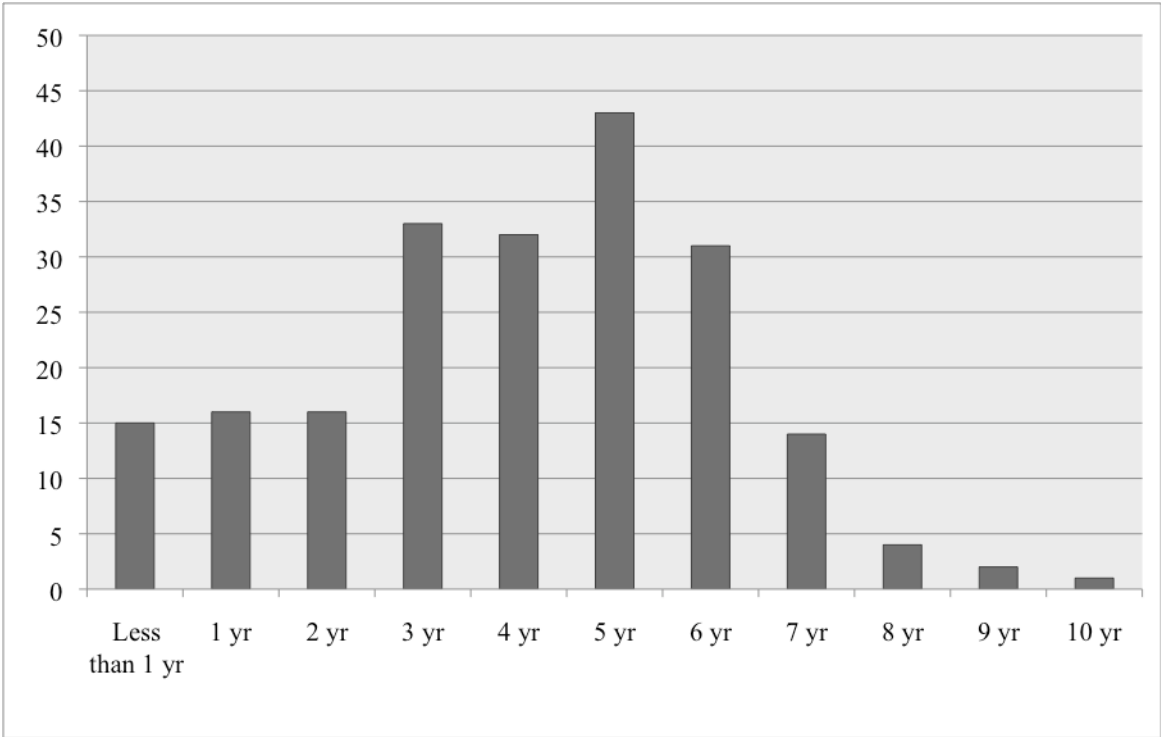
B.   How MusicBrainz Works

As discussed in the introduction, there are several ways that people can engage with MusicBrainz. This section outlines the technical process of using MusicBrainz to tag metadata; the editing and voting processes; communication channels and documentation; and the process through which users add to and amend style guidelines.

Anyone with access to the web may browse and search for data on the MusicBrainz website. Similarly, anyone may install a MusicBrainz-supported tagging application and look up information, scan tracks, and tag metadata. MusicBrainz supports several tagging applications, including Jaikoz, MagicMP3, and MusicBrainz Picard, which is the official MusicBrainz tagging application, written in Python. Both MagicMP3 and Jaikoz must be purchased, although both programs offer a free trial. Because Picard is the official MusicBrainz tagger, this section focuses on its functionality.

A user of Picard has the option to find metadata for untagged tracks in two ways: if partial metadata exists, such as track or album names, a user can choose "Look-Up," which will attempt to match the metadata against metadata existing in MusicBrainz; if the file has no metadata, a user can choose "Scan," which will generate a fingerprint[1] of the track and request information from a non-MusicBrainz server, MusicDNS, in an attempt to match the fingerprint to a portable unique identifier (PUID) in the database. A PUID can only be generated if music analysis of the full track has already been performed, and this analysis cannot be done with Picard "because the process is closed source and Picard is GPLed" [8]. The General Public License (GPL) dictates that any derivative work resulting from modification of the software must also be licensed under GPL and modifiable, and the source code be made available. In this case, that would also mean any source code relating to the proprietary music analysis process. But exposing code would violate MusicDNS' intellectual property rights over the music analysis technology—that is, people could figure out how it works and jeopardize their business model. The analysis instead must be done with a program called MusicIP Mixer. If MusicDNS has a match for the fingerprint from Picard, it sends the PUID back to Picard. Picard then looks for that PUID in the MusicBrainz database. If it finds the PUID, it returns it.

Because Picard cannot generate PUIDs, there is no guarantee that Picard will find the track in the database after a scan since there not be a PUID for it. In cases like this, user can run a search query for the item they wish to tag in Picard's search box, which will pull up a an indexed list of search results, ranked by likelihood of a match, on the website. A green "Tagger" button appears that, when clicked, sends the entity information to Picard (screenshots shown in Figure 7). A user can then drag the track from the list of unmatched tracks in the left-hand panel and drop it on the entity information in the right-hand panel. Users then save the tags, which adds the new metadata to the audio files.

---

[1] An acoustic fingerprint, also called an audio fingerprint, is a "content-based compact signature that summarizes an audio recording" [18:233]. They are created with technologies that extract acoustic characteristics, which are stored in databases like MusicDNS, and matched algorithmically to other fingerprints for the purpose of identification.

Figure 7: Screenshots of the Tagging Process

The following releases matched your query

Page [1] of 151 (Go!)   « < **1** 2 3 4 5 > »

Score	Release	Artist	Tracks	CD ids	Date	Type	Lang/script	Tagger
100	The American Song-Poem Anthology	Various Artists	28	1	us 2003	compilation	eng / latn	TAGGER
58	The American Song-Poem Christmas	Various Artists	21	0		compilation	eng / latn	TAGGER
17	American Folk Anthology Volume One	Various Artists	14	1		compilation	eng / latn	TAGGER
17	An American Folk Song Anthology	John Stewart (singer-songwriter (Kingston Trio))	12	1	us 1996	album	eng / latn	TAGGER

MusicBrainz Picard

Add Files Add Folder Save Submit PUIDs CD Lookup Cluster Lookup Scan » Album

Title	Lengtl	Artist
Track 11	2:34	
Track 12	2:27	
Track 13	2:20	
Track 14	2:53	
Track 15	2:12	
Track 16	2:23	
Track 17	3:39	
Track 18	2:22	
Track 19	2:11	
Track 20	4:07	
Track 21	2:21	
Track 22	1:56	
Track 23	2:35	
Track 24	2:32	
Track 25	2:37	
Track 26	2:20	
Track 27	2:15	
Track 28	1:41	
Clusters		

| Title	Lengtl	Artist
The American Song-Poem Anthology (0/28)	7...	Various Arti...
01 Do You Know the Difference Between Big Wo...	3:30	Gary Roberts...
02 Rat a Tat Tat, America	3:01	Dick Kent
03 I Like Yellow Things	2:14	Bobbi Blake
04 I'm Just the Other Woman (remake)	2:37	The MSR Sin...
05 Human Breakdown of Absurdity	4:20	Norm Burns ...
06 Maker of Smooth Music	2:19	Dick Kent
07 Beat of the Traps	2:39	Rodd & The ...
08 Richard Nixon	2:54	Rodd & The ...
09 Jimmy Carter Says 'Yes'	2:23	Gene Marshall
10 Convertibles and Headbands	4:05	The Music M...
11 How Long Are You Staying	2:34	Bill Joy
12 Little Rug Bug	2:27	Rodd Rogers...
13 I Lost My Girl to an Argentinian Cowboy	2:20	[unknown]
14 City's Hospital Patients	2:53	Teri Summer...
15 Ecstacy to Frenzy	2:12	Rodd Keith
16 All You Need Is a Fertile Mind	2:23	Gene Marshall
17 The Moon Men	3:39	John Muir
18 The Palace Roses	2:22	Todd Andrews

Original Metadata

Title:
Artist:
Album:
Track: Length: Date: - -   Lookup

New Metadata

Title:
Artist:
Album:
Track: Length: Date: - -   Lookup

Files: 28, Pending Files: 0

Hemerly 25
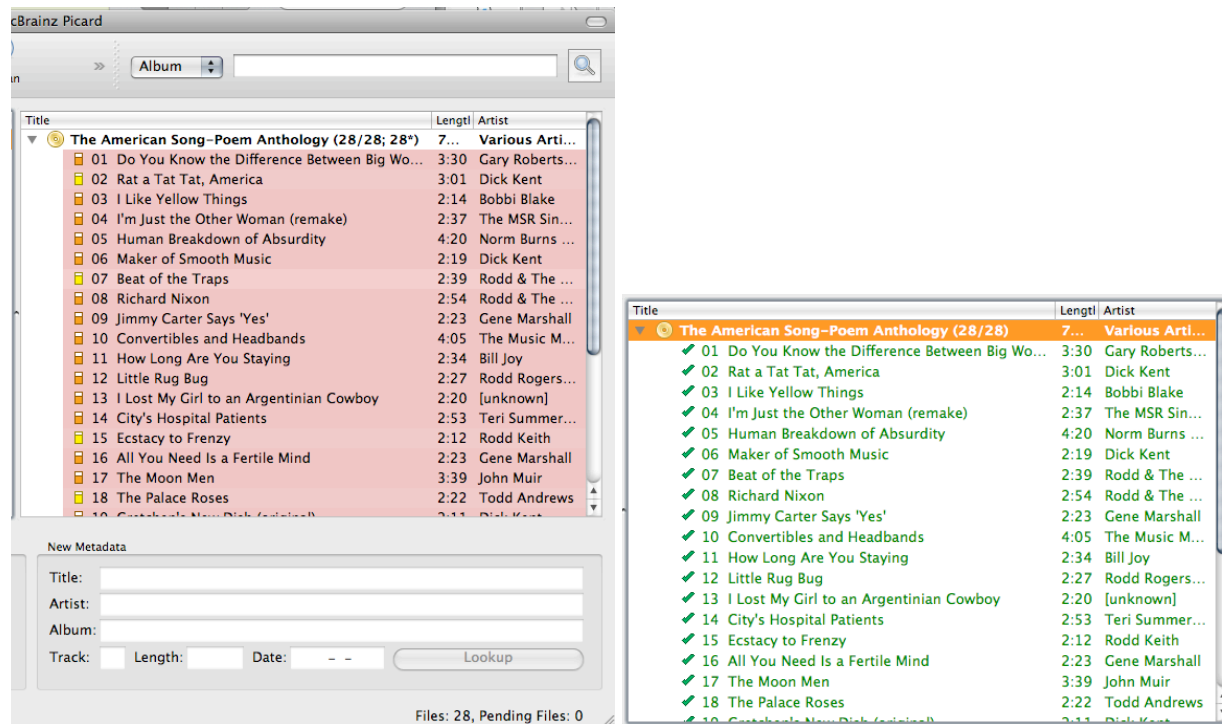
Dependence on a proprietary system for music lookup is one of the drawbacks of MusicBrainz, and they have already had to transition from one service to the current one because the first company went out of business. Because of this, one user has been working on his own version of MusicDNS called AcousticID, that will store the links between fingerprints and tracks to that, if MusicDNS no longer existed, MusicBrainz would at least have the recorded links between fingerprints and identifiers stored in an open system. The developer intends for it to serve as a replacement or MusicDNS, but notes, "I'm far from there yet" [28].

Because Picard prohibits quick fixes to metadata within the program, if something is not in MusicBrainz, Picard will prompt a user to add it to the commons. A user must add a release to MusicBrainz before he or she can fix tags with the tagger, and this is the most common way users in the study sample became editors. In order to edit the data within MusicBrainz, a user must register an account by creating a username and password. Once a user has registered with the site, he or she may begin editing the data and is officially an editor. Changing the data used to be called "moderating," but has been changed over the years to "editing" [9]. As of Monday, April 18, there were a total of 541,779 registered editors, 1337 of who edited the week prior (these stats, available at http://musicbrainz.org/show/stats, are updated regularly). In addition to its volunteer editors, MusicBrainz has, according to Kaye, "3.25" total paid employees: Kaye, two engineers/developers, and the quarter-timer, a community manager. All paid employees participated in the interview phase.

When an editor enters an edit, whether he or she adds a new entity to the database or corrects data that is already there, the entry form includes a field for Edit Notes (Figure 8). Users are encouraged via a message below the "Enter Edit" button to add a note, writing, "Please note that it is highly recommended to enter an edit note. If you do so, your edit is more likely to be voted on, and unanimously accepted by the other MusicBrainz editors." An Edit Note usually includes

a link or an informal citation to another source for music information like an artist's or label's website, an online music store, or, if nothing else is available, Wikipedia, Discogs or FreeDB. As one editor explained, "I usually have 5-7 tabs open, some of those tabs are from MusicBrainz from an artist, an album, from some other artist if it's like linking information, and from official artist homepage, where I can get accurate information, or Wikipedia if it has more complete information" [36]. Several interview subjects noted that Discogs is an especially good source for information about electronic music because people who self-release tend to add their releases to that site. Other users are wary of the quality but feel that "a citation from Discogs is better than no citation" [39].

Figure 8: Screenshot of the Edit Release Title form

Edit Release Title

- This form allows you to edit the ⓤ **title of this release**. Please read the ⓤ **style guidelines** before entering any edits.

Current Title: ⓒ Matador at 21 (disc 3: Days of Whiskey and Tears (1996–2001))

New Title: Matador at 21 (disc 3: Days of Whiskey and Tears (1996-2001)) ⓦ (Guess case)

Enter edit note (optional):

Please enter an edit note here. Thank you

(Cancel) (Enter edit »)

About edit notes: Please note that it is highly recommended to enter a **edit note**. If you do so, your edit is more likely to be **voted on, and unanimously accepted** by the other MusicBrainz editors. If your edit receives the required number of unanimous yes votes, it will be applied instantly, even before the voting period elapses in which edits usually are kept open. For detailed information on how long edits stay open and how many votes are required for it to pass unanimously, please see our edit conditions page.

As discussed in earlier in this paper, MusicBrainz does not include genre or style in the metadata information. That is left to the MusicBrainz user to add once she has tagged a file in her personal music collection. With the exception of artist annotations and tags, the data in MusicBrainz is factual information that can be pulled from liner notes, news articles, or other objective material. A plug-in exists for Picard that applies folksonomic tags to tracks, and one editor interviewed likes to push all of the tags into the genre field and choose the one he thinks is most appropriate [36]. Another editor does not use genre tags at all in his collection, choosing instead to rank his music by personal ratings he assigns through his music player, Amarok [27]. Different collection management habits dictate different metadata needs.

Documentation for MusicBrainz begins on the MusicBrainz wiki and is transcluded, or published, to the project's main online documentation as it is approved by Transclusion Editors. Transcluded documentation includes docs targeted at new users (e.g., "Beginners Guide" and "How Editing Works"); docs conveying the philosophy and social rules of MusicBrainz (e.g., "Code of Conduct," "MusicBrainz License," and "Social Contract"); and the style guidelines, which govern the standards for representation of data within MusicBrainz (e.g., "Official Style Guidelines," "Capitalization Style Guidelines," and "Classical Style Guides"). The MusicBrainz wiki also includes a number of pages that have not been formally transcluded into the site's

documentation, such as "History of the Style Council," which documents the changes in the processes that determine how metadata should be structured and represented. Each editor also receives his or her own wiki space, and users often use that space as a way to list their areas of musical expertise or editing interests.

The MusicBrainz discussion forums are a communication channel in which users can post questions for other editors or general comments on the site's structure. The forums are divided into two major categories: "Support," with the subcategories "Website Support" and "Tagger Support"; and "Discussion," with the subcategories "Feedback and Ideas," "Style Discussion," and "General Discussion." For new users, the forums serve as a place to clarify issues raised in edit notes or to get additional support on how to follow the site's style guidelines [44], which can be daunting to a new user. Established editors answer new users' questions and engage in discussions about data representation, special cases, and style guidelines, or suggest improvements or general desires.

MusicBrainz editors also communicate outside of the MusicBrainz website through a number of channels. There are three main email mailing lists: Users' Mailing List (general discussions), Style Mailing List (style guidelines-specific discussions), and Developers' Mailing List (issues that affect the project's technical developers). There are also several other "low traffic" email lists, including the Data-Feed Mailing List and the Italian Mailing List, but the main lists are where most of the list communication happens.

Finally, the community uses Internet relay chat (IRC) for real-time communication. There are two channels: #musicbrainz, which is a general discussion room for editors, and #musicbrainz-devel, a channel more specifically used by MusicBrainz developers for technical discussions and weekly developer meetings, which are held every Monday. MusicBrainz makes the chat logs from IRC channels available on the website (http://chatlogs.musicbrainz.org) so that everyone can see what discussions have taken place. The discussions range from technical and style questions to more general discussions about music. Editors also occasionally engage in friendly, non-music-related conversations as well, which will be discussed in more detail in a later section.

Representation of the data, down to hyphenation and capitalization, is governed by community-established guidelines. The guidelines are intended to be flexible suggestions to guide users in things like capitalization, punctuation, and representation, but editors take them quite seriously and many feel that the guidelines are strict rules only to be bent in special cases. As a result, the guidelines are more like community-established standards than flexible suggestions. Editors who violate the guidelines without having proper evidence that an artist intended it as such face the possibility of having their edits rejected by the community.

An edit is committed to the database as soon as it receives three unanimous "Yes" votes, and fails when it receives three unanimous "No" votes. Two "Yes" votes and a "No" vote will hold the edit in the queue. Editors can also choose to "Abstain," which is a public way of declaring, "I don't know." If an edit of default quality receives zero votes, the edit will be automatically accepted after 14 days (the majority of the data in MusicBrainz is "default quality"). Editors are not able to vote on other edits until they have had 11 edits accepted and have been members for at least two weeks, with a confirmed email address. The voting mechanism gives editors time to

review changes to the database and more quickly approve those things which are deemed correct and reject those things which are incorrect or do not conform to the style guidelines.

Debate about major changes to the style guidelines occurs via the Style Council mailing list. Any editor can be a member of the Style Council by simply signing up for the mailing list and joining the conversation. An editor proposes a change by sending out a request for comments (RFC) to the mailing list, to which other "council members" then provide comments and feedback. As one interviewee explained, if the comments are mostly negative, the submitter must either make suggested changes or withdraw the RFC altogether. If the comments are mostly positive, however, then he RFC moves onto the next stage, which is Request for Veto (RFV). Everyone should have made any comments they were going to make on the RFC, but at the RFV stage, any Style Council member can also veto the change. If a council member vetoes it, the proposal basically goes back to "square one," as one puts it [33]. If nobody vetoes, the change is approved.

C.   Patterns & Processes

Patterns of Contribution

This section details the ways personal music collection management, musical taste, and free time affect editors' contributions to the project. Among the patterns identified is a notable disparity between editing and voting on edits as well as a positive correlation between artist subscriptions and edits entered by an editor. Editors tend to contribute in waves linked to their free time and acquisition of new music.

Survey respondents were provided a list of 16 activities that registered editors can perform in MusicBrainz and asked to select "Yes," "No," or "Uncertain" for each option based on whether he or she has performed the activity. The results are shown in Table 4, ranked by total of "Yes" responses. The number one activity on MusicBrainz performed by respondents is "Adding releases to MusicBrainz." As noted in the previous section, finding that something in a person's music collection is not yet in MusicBrainz through the tagging process is an important point of entry for new editors as well as a way existing editors stay engaged. That is, when attempting to tag a new addition to one's collection, an editor who finds it does not yet exist in the database will be prompted to add it.

Table 4: Which of the following activities have you performed?

Activity	Yes	No	No Answer
Adding releases to MusicBrainz	229	9	1
Relating entities to non-MusicBrainz URLs	214	15	7
Vote on changes made to your artist subscriptions	187	46	4
Subscribe to new artists	176	59	3
Add annotations	173	38	27
Make notes on edits and votes	168	54	16
Vote on changes made by your editor subscriptions	134	84	13
Subscribe to editors	131	101	6
Participate in forum discussions or mailing lists	120	108	10
Editing info relating to MusicBrainz entities on Wikipedia, LyricWiki, or Discogs	114	101	22
Adding/tracking/editing bugs	82	143	14
Editing documentation or informational pages on the MusicBrainz wiki	66	163	9
Working on a smaller community-defined project	57	164	17
Maintain a MusicBrainz server	21	231	2
Contribute code	19	211	6

This suggests that acquiring new music is linked to activities in MusicBrainz, and the interviews support this pattern. Several editors described their process when acquiring new music or importing music from physical media artifacts, and included tagging with MusicBrainz and adding an absent release as a crucial step in their managing music collections. One editor mentioned that he actually removes all existing tags and applies the MusicBrainz tags to the music, which requires that, because Picard prevents on-the-fly tagging, any new album must be added to MusicBrainz before he can properly tag it. He says, "I like the idea of having URIs for all of my music, so if it's not in MusicBrainz, often it's sufficiently obscure that I feel like if I don't put it in then it's probably not going to end up there any time soon, anyway" [37]. Metadata consistency is extremely important to him.

One editor, a recreational DJ, has made sure that all of his music is in MusicBrainz because he has slowly been working to convert vinyl to MP3, and having it in MusicBrainz not only makes the metadata available to others, but because of frequent data loss, it also makes his life significantly easier: "When I actually get around to ripping my vinyl to MP3s and I want to tag it, to have it done once is more convenient than having to retag it over and over and over again, because I end up losing my music somehow." [35]. For him, it is an information need turned necessity—an extrinsic motivation like open source software contributors who contribute because they want to use the software [55].

But because contribution is so closely linked to acquisition, and because not all editors acquire new music regularly, many contributors described waves of contribution that wax and wane based on how often they get new music. One user told me, "I still do editing but I only need to edit when I add something to my collection and that's slower. But at the start there was a greater amount of activity from me, and I also had a lot more time" [27].

A change in the amount of free time a user has was a common justification for a decrease in contribution. "It's mostly things in my life changing which caused me to not have time for it anymore" [31] said one editor, and several others simply mentioned "not enough time" as a reason for their waning contribution. Another editor whose contribution has been greatly affected by lack of free time talked about the difficulty in shifting from heavy involvement to lower-level involvement: "The burnout curve as you might call it or something like that — when you're that involved with something it's hard to shift back to a low level engagement without it really just kind of dropping out almost entirely" [39]. MusicBrainz continuing to grow and thrive as an active community, then, depends on a constant influx of new users finding the site, passing their collections through taggers, and signing up to contribute.

One task that suffers due to time constraints is voting. There is a large gap between the number of edits made to MusicBrainz and the number of votes made on those edits. A theme that emerged in interviews is that editing is qualitatively easier because you can simply link to a source online that justifies the edit. Editors are hesitant to vote unless they really know about something, and prefer spending their time confirming what they already know and adding data to fact-checking the contributions of editors.

> "To make edits I feel like there's usually enough resources on the Internet to just find something and put something out there so that it exists in MusicBrainz [...] I can find a tracklisting and I can put it up and I feel comfortable doing that. I don't necessarily feel comfortable voting unless I have knowledge because then I have to be able to say 'Yes this is true' or 'Yes this isn't' rather than just 'Yes I found this on the Internet, I think it's right' [37].

One of the survey's Likert-type questions asked respondents to agree with the statement, "I vote on edits only when I feel knowledgeable about the entity being edited" (see Table 5). Most respondents answered positively to this question, with the mean slightly above 3, which was "Agree" ($\mu = 3.16895$). This positive sentiment coincides with what came out of interviews.

Table 5: Likert-type Questions on Voting

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
I vote on edits only when I feel knowledgeable about the entity being edited.	3.16895	76	107	33	3	15

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

On January 15, 2011, one of MusicBrainz's employees provided statistics on all-time edits and votes within MusicBrainz. The results of his server query are shown in Table 6.

Table 6: Edits vs. Votes (all-time, as of January 15, 2011)

Edits	N	Percentage
Applied	10,268,963	97.4%
Deleted	137,045	1.3%
Failed due to voting	135,398	1.28%
Total	10,541,406	
Votes		
Yes	4,241,327	55.8%
Abstain	3,005,106	39.5%
No	350,732	4.6%
Total	7,597,165	

Because an edit generally requires three votes to either fail or be accepted, it would reason that the total number of votes over time would be significantly higher than the number of edits applied. Auto-Editors, whose votes are automatically accepted, are an exception to this, and as I will show later, have a higher mean of edits than non-Auto-Editors. However, the numbers still indicate that editors spend more of their time editing than they do voting. But this does not mean that the data quality suffers or that the voting process is pointless. Rather, the small percentage of "No" votes among total votes indicates that MusicBrainz editors tend to make solid contributions and the project does not suffer from trolls who purposely set out to vandalize the database.

The gap between editing and voting could exist for a number of other reasons: many editors indicated that they do not vote unless they feel knowledgeable, and as discussed in a later section, they tend to work within niche areas of music in which they have knowledge, often because of their personal listening habits. It could also link to the time issue, where time is better spent building the database by editing than it is policing the database by voting. Because voting does not directly expand the knowledge in the database, voting may be seen as an activity secondary to satisfying an information need, thus falling outside the scope of the need-based extrinsic motivation. However, it may also just be as simple as voting not being enjoyable, hence lacking the intrinsic enjoyment-based motivation that exists for editing. As one editor said, "I guess a part of it's just 'cuz voting is not particularly fun" [35]. Among the editors interviewed, voting is definitely not a priority.

To make it easier for editors to vote on edits to entities with which they are familiar, MusicBrainz allows editors to subscribe to artists and receive a daily email digest of changes made to their artist subscriptions. However, while interviewees subscribe to artists, they frequently ignore the subscription digest or send it straight to the trash. One editor saves the emails without reading them, but occasionally goes back to look at them when he has time. But one drawback of the disparity between voting and editing is that voting allows editors to make notes on other editors' edits, and these notes are valuable to new editors trying to learn the ropes. Not only are the style guidelines thorough and complex, some accepted practices that develop over time are never codified in the style guidelines. As one user told me, "a few prolific editors doing it in a particular way and that kind of becomes the norm" [30].

While there exists no significant relationship between the number of artists to which an editor in my sample has subscribed and the number of votes he or she has entered, there is a very significant relationship between the number of artist subscriptions an editor has and the number of edits he or she has entered. A pair-wise correlation test returned a correlation coefficient r of 0.9024, indicating a nearly linear relationship between edits entered and artist subscriptions. With a p-value of 0.0000, this is a highly significant correlation. Subscriptions may not actually encourage more people to vote, but there is a very strong positive relationship between subscriptions and edits entered.

Peer-produced Information Science

In the process of developing guidelines and structure for metadata in the commons, MusicBrainz editors engage in peer-produced information science by participating in "the design of metadata structures" for MusicBrainz, an information pool that ultimately aids Music Information Retrieval (MIR) [71]. As the National Information Standards Organization (NISO) notes, "Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content" [62:10]. The body of core editors and Style Council members who regularly participate in discussions about structure in essence create controlled vocabularies for metadata, and function much as NISO would in bringing together multiple voices to establish metadata standards. This section discusses the architecture of the metadata structure from an information science standpoint, and shows a split in the community regarding whether or not an editor should own a copy of a release they add to the database.

There exists a strong connection between the desire to organize and the need to retrieve—that is, consistent metadata standards make it easier to find things. For all interviewed editors, proper metadata is crucial to finding music to listen to: "If the files are properly tagged it makes it much easier to search for things in your media player" [31]. One user compared the willingness to sit down and enter detailed information about music to Wikipedia's strength when it comes to information about current events: "Music is sort of always a current event in that way, where it's always something that there's someone who's interested enough to just sit down and figure out what little bits of data need to go in the right fields so that it works" [37].

MusicBrainz has developed a system of disambiguation to overcome the metadata problem of *polysemy*. Polysemy occurs when one word, sign, or symbol has multiple meanings. In the case of music, multiple artists may share the same name. For example, MusicBrainz has three artists with the artist name "Death." In order to distinguish between these three artists, and any future artists of the same name, editors use the comment field to enter a description that distinguishes the Deaths from one another. The three artists called Death are "the death metal band of Chuck Schuldiner," "German Techno producer Thomas P. Heckmann," and "70s punk/hard-rock trio (brothers Hackney) from Detroit." Artists are also identified as "Person," for an individual artist, or "Group," further aiding differentiation. There are six artists named "John Smith" in MusicBrainz, with disambiguation comments similar to those for the bands called Death.

All but one of the interview subjects mentioned specific challenges that classical music pose to the structure of metadata in MusicBrainz. However, this is not unique to MusicBrainz; rather, it

is a problem faced by all information professionals tasked with designing organizing systems that need to accommodate music and musical entities. Smiraglia explains:

> Collections of musical documents are unique among collections of documents (in libraries, bibliographies, etc.) in that the influence of repertory is such that a given collection will have many instantiations of the same musical work—one Tchaikovsky Fifth Symphony (a musical work), but a dozen scores of different sizes and formats, and dozens of recordings, not to mention excerpts and arrangements. [71]

For example, Antonin Dvořák's Symphony No. 9 in E Minor often includes "From the New World" in the title of a recording or is simply referred to as "New World Symphony." Dvořák's page in MusicBrainz shows many different versions of the same composition with different names, performed by different bodies, often recorded in the same performance as other works by the same or different composers. Popular music experiences the same issues when it comes to covers—an artist performing the work made famous by another artist—and appearances of the same song on multiple releases, both live and in-studio.

The forthcoming Next Generation Schema (NGS) includes a number of structural changes that will not solve the classical problem, but are steps in the right direction, notably the distinction between work and recording. This distinction echoes the International Association of Library Associations and Institutions' (IFLA) Fundamental Requirements for Bibliographic Records (FRBR), which outlines levels of abstraction in bibliographic entities:

- "work," the "distinct artistic or intellectual creation;"
- "expression," the form a work takes every time it is "realized";
- "manifestation," the physical (or digital) "embodiment of a expression";
- and the "item," a single instance of a manifestation [47].

For example, in the case of Dvořák, "Symphony No. 9 in E Minor, 'From the New World'" would be the "work," the composition created by Dvořák; a 2010 performance by the New York Philharmonic would be an "expression"; a CD released by Sony Classical is a "manifestation"; and the copy of this CD in my CD collection would be an "item."

The distinction in the NGS is not quite as detailed as FRBR's abstraction, but the elements are there. NGS includes "work" and "recording," where the work will be "Symphony No. 9 in E Minor, 'From the New World'" as above, and the word "recording" stands in for "expression." MusicBrainz already includes fields for different manifestations of releases, called Release Events, where editors can fill in separate events for releases in different formats, like vinyl, digital, cassette, and CD. The item, then, is the digital file in a user's collection, tagged with the metadata that represents the various levels of abstraction.

This structural change allows for a single work to exist as its own entity within the database instead of existing as a number of different recordings of the same thing. Smiraglia notes, "In a digital era of music information retrieval, the degree to which differing sonic instantiations

represent the same work has an epistemological basis" [71:754]. For example, Dvořák should never be a *recording artist* because he never recorded his own work. But whether the recording artist credit is the Royal Philharmonic Orchestra or the New York Philharmonic can also itself convey cultural and contextual information about directors, featured soloists, and ensembles that is indispensible to a music fan. This structure will also work well for entities in popular music, such as a cover version of a song, an album reissue, a compilation, or a recording of a live performance.

While a link between collecting music and editing exists, editors are encouraged to make a variety of contributions that will improve the site's metadata. Editors use multiple sources to justify edits in notes, from other open sites like Wikipedia and Discogs, to official artist pages and label sites, but interviewees do vary on what is a "reputable" source. For example, the electronic music fans I spoke with use Discogs frequently, but others have concerns about the quality of the data. The prevailing sentiment among interview subjects is that data can be imported from FreeDB but FreeDB should never be used to justify an edit—"official" sources are best, when available.

Survey respondents showed mixed feelings in response to the Likert-type question, "An editor should only add a release if s/he owns a physical or digital copy" (see Table 7), which shows that, while the majority of respondents to this question selected "Disagree," 25% of respondents (N=233) responded positively ($\mu = 2.0987$). This indicates a subgroup of editors who seem to believe that people should not add things to the database that they do not own, but that the majority of editors believe owning the music is not necessary to make a contribution.

Table 7: Information Source Likert-type Question

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
An editor should only add a release if s/he owns a physical or a digital copy.	2.098712	17	40	125	51	2

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Distributed Work and Musical Taste

In a 1956 paper titled "The Social Nature of Musical Taste," sociologist John H. Mueller wrote, "We cannot evade the conclusion that the answer to our question, 'What is good music?,' may differ from culture to culture, from epoch to epoch, from person to person, and even from time to time within the same person, as well as the particular occasion" [61:120]. This variance in musical taste is evident within MusicBrainz, in the relationship between editor contributions and what they like to listen to, and the growth of the database depends on it. There is also a difference between editing habits that increase the depth of the database and editing habits that contribute to the breadth.

As discussed in section "
Collective Action and Constructed Cultural Commons," for a constructed cultural commons like MusicBrainz to exist, information must necessarily be created before it can be used [56:672]. In

the case of MusicBrainz, this information is metadata about music, including artists, releases, and advanced relationships between database entities. Due to the practical application of the data—tagging musical releases—new music is often added a release at a time. Thus, the amount of information available depends on the heterogeneity of editors' musical taste—that is, more editors with varying taste means a more releases spanning more types of music. Several subjects noted that when it comes to "popular music," entries are generally complete, so adding releases in "eclectic," "niche," or "obscure" areas—all terms used by interview subjects to describe their taste in music—helps to increase musical coverage in the database.

Figure 9 shows the results from two Likert-type questions relating to musical taste and knowledge. In response to the question, "Entities I edit reflect my taste in music," the mean (μ = 3.2217) indicates positive sentiment, supporting a connection between musical taste and editing. As the responses to "Participating in MusicBrainz allows me to share my music knowledge with others" indicate, participation is a social platform through which users demonstrate and share what they know about music. Sharing this musical knowledge provides both an enjoyment-based and an obligation/community-based intrinsic motivation [55].

Figure 9: Likert-type Questions Regarding Musical Taste and Knowledge

	Mean	**Strongly Agree**	**Agree**	**Disagree**	**Strongly Disagree**	**No Answer**
Entities I edit reflect my taste in music.	3.2217	100	94	4	3	2
Participating in MusicBrainz allows me to share my music knowledge with others.	3.3463	104	101	22	2	4

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Editors can also contribute to the metadata commons in ways other than adding nonexistent entries. Details that are not necessarily fields one might find in a music player, such as group start and end date, Amazon ID, and Advanced Relationships (see How MusicBrainz Works) are also data points that editors can contribute. One editor I spoke with discussed a particularly interesting resource for information: obituaries. When an artist dies, this editor takes advantage of the artist's presence in the news and fills in the information for the dead artist in MusicBrainz. And he's not alone: there are "a couple other people who keep an eye on the obituaries, and it's actually when I heard that other people did that that I started doing that" [34].

The work done by MusicBrainz editors can be characterized along two axes:

(1) **Breadth:** Building outward with new artist and release entries to cover a multiplicity of musical styles and artists, largely dependent on passionate fans who like specific "niche" styles. MusicBrainz's breadth depends on editors with knowledge of styles outside of the range of "popular" music. An editor who adds new releases in a niche style like Japanese pop (J-Pop) or an editor who adds entities in a certain style of death metal increase breadth. Taste definitely drives contribution: "Occasionally there are people entering or fixing data for other reasons, but usually it starts with someone, somewhere bringing it up because they wanted to listen to that album" [31].

(2) **Depth:** Building upward by fixing existing entries, expanding relationships and links, and adding details to existing entities. MusicBrainz's depth depends on editors who focus on filling in missing fields and making corrections to existing data. Adding death dates from obituaries, links to Wikipedia pages, and album art are ways that editors increase the depth of the database. One editor said, "You can get to this level of edits that's pretty much not useful to anyone and the nice thing is that they are probably useful to someone, which is what's always weird about it" [37].

Users describe their taste in varying ways, from more traditional genre and style distinctions—drum & bass, metal, electronic—to terms that describe the music's sound. One user likes things that are "jolting" or "surprising," [32] while another likes things that are "heavy and fast" [36]. Regional location also serves as a niche-defining factor, at least when it comes to editing. One user used his region as a way to begin editing data, stating, "I'm from Slovakia, so I started editing local artists/albums. There weren't many editors from this area of Europe, so Slovak/Czech/Polish albums were my main target" [28]. While he was not always a fan of the artists he entered, he was able to read websites that "nobody else could" because of his native language. Slovakian music became his "niche."

Mueller argues that musical taste is deeply embedded in social life, and musical practice itself is a societal institution. He writes, "They are the beneficiaries of an esthetic conscience-analogous to moral conscience—which labels discrepant tastes as wrong, and resists radical intrusion of new codes and systems of taste" [61]. Because of this deeply ingrained musical conscience, we often see conflict where different tastes collide. However, within MusicBrainz, editors with varying taste respect each other, "which you don't always see in people with completely different tastes in music" [39].

It is because of the aforementioned depth and breadth issues that the community welcomes editors of all musical taste. Without people liking things that are unusual and unpopular, the database would be homogenous and useful to far fewer people. As Margaret Mead wrote in the 1972 article "Music Is A Human Need," "…No human culture has ever been built which was liked equally and in the same way by every person who has grown up learning that way of living and seeing the world. It is just these differences in individual taste that make for variety and change" [60:27].

Editors also develop bonds with other editors who share their taste and work in their niche areas. Says one editor, "In a lot of cases there's less overlap in the stuff that they're editing so they don't necessarily care about what the other people do but there's an attempt to harmonize and cooperate" [39]. Another interview subject, struggling with representation of French music, found help "just by surfing around" [38]: "I saw that this guy was systematically involved in editing content or tracks from French artists so I figured that they were first, probably francophone, and then that they probably knew a lot about the things that were bothering me." Thus, shared musical taste and interest also creates smaller clusters of users with similar interests who work together to expand breadth and depth in a niche.

Discovering MusicBrainz

Learning how MusicBrainz editors discover the project is one way to understand how it relates to the music information landscape at large and to individuals' information needs. While editors come to the project from a number of sources, those respondents who learned about MusicBrainz through Last.fm make more contributions than those who did not come from Last.fm.

"How did you learn about MusicBrainz?" was a mutually exclusive categorical survey question where respondents were asked to select one of a source from a list of sources (Table 8). The most common way a respondent found MusicBrainz was during a web search for a way to clean up music metadata. This demonstrates that many MusicBrainz editors come to the site because of a very specific information need—that is, consistent metadata for their digital music collections.

Table 8: How did you learn about MusicBrainz?

Source	Frequency	Percentage
Web search to clean up my music metadata	83	35.93%
Online discussion group or forum	49	21.21%
Other	32	13.85%
Audioscrobbler/Last.fm	32	13.85%
Friend	13	5.63%
Blog or Twitter	11	4.76%
Can't remember	9	3.90%
Creative Commons	2	0.87%
TOTAL	231	

One of the question options was "Other," which allowed a respondent to enter open text into the field explaining how they discovered MusicBrainz. Thirty-two respondents entered either "Last.fm" or "Audioscrobbler," now a part of Last.fm, in the box. Because this was such a significant number of users, the variable was manually recoded in the .CSV to include Audioscrobbler /Last.fm as a coded category. Those who entered "Can't remember" were also manually recoded, and dummy variables were generated for each of the discovery sources.

Because Last.fm is an online radio/recommendation service dependent on good metadata for building a user's library of listening habits, the Last.fm service recommended that its users go to MusicBrainz to properly tag files. Last.fm was a frequent topic of conversation in interviews, and half of the interview subjects have Last.fm accounts. Additionally, Last.fm is a customer of the MusicBrainz live data feed.

Finding MusicBrainz Hypothesis: The relationship between Last.fm and MusicBrainz is an important one in terms of those who actively edit data and subscribe to artists. Editors who discovered MusicBrainz through Last.fm will have a higher mean of subscriptions and edits entered than those who did not. Other methods of discovering MusicBrainz will not show a significant difference.

Table 9: t-test Results: Edits entered based on responses to "How did you learn about MusicBrainz?"

	Respondents who said Yes		Respondents who said No			
	N	Mean (SD) log of entered edits	N	Mean (SD) log of entered edits	*t*	Cohen's *d*
Friend	6	6.132633 (0.6580)	160**	7.354181 (2.3156)	3.7541+	0.7175
Blog/Twitter	10	6.824719 (3.2754)	156	7.341138 (2.2205)	0.4914+	
Online discussion forum	36	7.292926 (2.0014)	130	7.314765 (2.3676)	0.0505	
Last.fm	28***	8.429282 (2.2129)	138	7.082934 (2.2425)	-2.9029	0.6044
Web search	55	7.04716 (2.2424)	111	7.440279 (2.3083)	1.0425	
Other	21	7.199127 (2.5698)	145	7.32609 (2.2530)	0.2370	
Can't remember	9	7.15812 (2.1832)	157	7.318737 (2.2996)	0.2043	

N = 166; diff = mean(no(0)) – mean(yes(1)); +SD inequality required Welch's unequal t-test
** pr(T > t) = 0.0012; *** pr(T<t) = 0.0021

Table 9 shows the results of t-tests performed on the mean of the log of edits entered for those who did and did not find MusicBrainz through each of the listed sources. The log of the number of edits entered was used because the *entered* variable was not normally distributed. Tests of standard deviation equality were conducted before running *t*-tests in order to determine inequality so that the necessary Welch's *t*-tests could be run on those with unequal variance. Only one respondent chose "Creative Commons," therefore a *t*-test could not be performed on that variable and it has been left out of the table.

The results show that there is a positive association between users who discovered MusicBrainz through Last.fm and editing activity, and with a *p*-value of 0.0021, this is of high statistical significance. The null hypothesis is rejected and the one-tailed "Finding MusicBrainz" hypothesis is statistically supported. Additionally, with a Cohen's *d* of 0.6044, there is a large effect size [21], indicating strong practical significance.

The results also show that there exists a negative association between editors who discovered MusicBrainz via a friend. That is, the *logentered* (log of total edits entered) mean of those respondents is actually lower than those who did not discover MusicBrainz via a friend. The *p*-value of 0.0012 indicates high statistical significance, and a Cohen's *d* of 0.7175 shows a large effect size, indicating very strong practical significance. Thus, users who come to MusicBrainz because of a friend's recommendation do not tend to be high contributors. But with only an N of 6 respondents who found MusicBrainz via a friend, it is too small of a number from which to draw a reliable conclusion about whether or not this may be generalized to a larger sample of editors.

Music Discovery through Contribution

In the paper "Birds of a Feather Sing Together," Mark proposed a theory of musical preference that suggested niche patterns develop because of "homophilous network ties," and notes the time and energy required to develop a preference for music as a reason why people often do not develop a preference for all types of music. Someone who has more free time will be more likely to develop a taste for music to which she has been exposed than someone who faces "many time constraints" [58:456]. While he was speaking generally about the affect of one's exposure to others' musical preferences on one's taste, MusicBrainz editors have found that more time spent voting and editing leads to music discovery, and those who have discovered new music through MusicBrainz have contributed more edits to the database than those who have not discovered a new artist. Editors who contribute more of their time to MusicBrainz are likely to discover more types of music.

Question 3 on the survey asked, "Have you ever discovered an artist through the MusicBrainz database?" Of the 238 respondents who answered this question, 87, or 36% chose "Yes." Answering "Yes" presented the respondent an open-ended text entry box asking them, "Briefly, how did you make this discovery? Please describe." One survey respondent wrote, "I often discover new artists while making edits to the MB database, such as when I'm correcting spelling/capitalization in cover songs, or when disambiguating between artists with the same name." Another respondent leverages the Advanced Relationships to discover new artists, writing, "You see that other artists collaborated with the artist you were looking at, and then discover them."

Despite the previously discussed disparity between editing and voting, a few editors in the survey sample do vote regularly and often discover new artists by voting. The process of verifying the accuracy of an edit becomes a method of discovery. "I have usually discovered new artists on MusicBrainz by voting on open edits for artists," wrote one editor on the survey. "Sometimes I am curious about an artist name and click on it. I will then follow links that exist in the artist relationships, or Google for more information and then link relationships myself if they are not present." Editors are exposed to "random artists when voting on edits," and their curiosity these editors to find out more about artists.

But how the editors interpret the word "discovered" varies. Some editors actually listen to the music and add it to their collections. One editor wrote, "I spot names which sound interesting while editing and voting and go find out what the music sounds like." Another editor wrote, "Found a messy set of albums for a band, downloaded some stuff to try and fix/listen to—ended up buying almost everything I could find by them."

Other editors just interpreted "discovered" to mean read about or were informed of. One editor explained in an interview, "One of the things I liked about MusicBrainz was that it actually exposed me to a bunch of different kind of musical artists and genres, at least in the sense of reading about them, not necessarily actually listening to them" [39]. The same user occasionally sought the artist or track for download but "less common" than just reading about it. The Slovakian user mentioned in the previous section started with music he listened to, but in relation to the amount of data he has contributed in his area, "I didn't even know about the most of it but it was interesting to read about older bands and put the knowledge to MusicBrainz" [28].

Whether new knowledge maps to musical preference remains unknown, but it is a potential area for follow-up. However, survey respondents overwhelmingly agree with the statement, "MusicBrainz has helped me expand my knowledge about music," shown in Table 10, with a very positive mean ($\mu = 3.2217$) on the Likert-type scale. This demonstrates that MusicBrainz editors have found the database useful in at least learning about new music.

Table 10: MusicBrainz has helped me expand my knowledge about music.

	Mean	**Strongly Agree**	**Agree**	**Disagree**	**Strongly Disagree**	**No Answer**
MusicBrainz has helped me expand my knowledge about music.	3.3169	87	108	34	3	3

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

As Mark concluded in his paper, "As is the case with all activities, participation in musical activities requires time and energy" [58:455]. To the extent that editing and reading MusicBrainz is a musical activity, people spend varying amounts of time editing data. Those who have entered more edits have spent more time contributing to the commons, thus more time engaging in this musical activity. This leads to the next hypothesis.

Discovery Hypothesis: Because of the patterns of exposure, editors who have discovered an artist through MusicBrainz are likely to have entered on average more edits—spending more time with the database—than those who have not discovered an artist through MusicBrainz.

Table 11: t-test Results — Have you ever discovered an artist through MusicBrainz?

	R selected Yes		**R selected No**			
	N	Mean (SD)	N	Mean (SD)	*t*	Cohen's *d*
Have you discovered an artist through MusicBrainz?	72	8.562961 (1.8384)	99	6.268651 (2.0947)	7.4396	1.16418

diff = mean(yes) – mean(no); pr(T>t) = 0.0000

The results of a *t*-test comparing the means of those who answered "Yes" and those who answered "No" on the variable *logentered* (log of edits entered) are shown in Table 11. The null hypothesis was rejected in favor of the one-tailed Discovery Hypothesis. With a *p*-value of 0.0000, this shows extremely high statistical significance and is not due to chance alone, and a Cohen's *d* of 1.1648 is a very large effect size, showing extremely high practical significance. Thus, those who have discovered an artist through MusicBrainz have made more edits on average than those who have not. The more time an editor spends editing the data, the more likely he or she is to discover an artist through MusicBrainz.

Consensus and Resolution

As covered in the "How MusicBrainz Works" section, contributors to the commons have developed processes and use several communication channels for community decision-making. However, interview subjects have varying opinions on how effective these processes are. Editors

split on the efficacy of the style discussions, citing arguments on the Style Mailing list as counterproductive. Editors also suggested that conflict can occur in edit notes, and that IRC is a preferred mode of communication for efficiency.

Nearly every interviewee mentioned the Style Mailing list, the channel through which changes are made to the style guidelines, and many editors put the mailing list discussions in a negative light. "I don't know if you've followed the Style Discussions" one editor commented, "but there's some contention on some really trivial things. People just like to argue" [32]. Another editor mentioned that these arguments are often about "really obscure things that debatably don't matter at all" [37]. Yet another editor avoids the style list "like the plague," saying that of all the mailing lists, "that is the one where people just go back and forth and back and forth constantly and it doesn't ever seem that a decision's made there" [35]. A "big blowup" caused a fourth editor to quit not only the style list, but also to resign from all lists entirely [30].

But not all editors see the style discussions as a waste of time. One editor sees it as an important way to constantly improve the quality of the database and to make the most of contributors' time: "I'd say it's important to get some of these things right, because an awful lot of work will get spent, and potentially wasted, if you don't have a strategy for these kinds of classifications and these conventions" [33]. For example, a recent style discussion [10] focused on the representation of "sort name" in the database, which is the name by which an artist is represented in an Indexed Search listing. Editors often use musical examples to demonstrate the implications of a particular decision, and in this case, on the mailing list on March 31, 2011, Les Claypool and his various musical projects served as the example (see Appendix). To some editors, debating the format of the sort name may appear to be a pointless exercise. But the editor who proposed this change argues, "sort names make it possible for external applications to create logical, alphabetical indexes" [10].

Sometimes, a proposed change splits the council members in such a way that it becomes "binary," and the opposing sides fail to reach agreement. One example of such a failure relates to composer and lyricist duos like Rodgers & Hammerstein or Gilbert & Sullivan. Most people commonly think of the artists of record as the duo, not just one of the pair, like Hammerstein without Rodgers or Gilbert without Sullivan. But because of the way the style guidelines have been laid out, the composer is the one who, technically, should be credited as the artist. As one editor explained, users faced off on this issue, and consensus was never reached on a proposed change that would have codified how this kind of pairing is handled in the database. Instead, there are a number of exceptions for special cases like the ones mentioned here, which users then have to look up in the style guidelines if they want to know how to handle things.

Because of the participation of the paid employees and the core editors, and because of its advantage of real-time communication, decisions are also made in IRC. A recent issue involving data quality in the NGS could have been negotiated over email, but because of the real-time way in which editors could suggest examples, the issue was negotiated in IRC "in a couple hours" [35]. And occasionally in IRC, an absence of disagreement or objection ends in a unanimous decision. For example, on February 23, editor CatCat posted in the IRC channel that the Cleanup of the Month project, a focused project to clean the data for a specific artist or area, would be

Bruce Springsteen: "18:13:46 <CatCat> since no one has argued, Bruce Springsteen is the new cotm [cleanup of the month]" (see Appendix).

In order to alleviate some of the difficulties in the decision-making process, Kaye recently hired a dedicated contributor to help manage and improve community processes, including the Style Council. Interviewed editors expressed respect for this editor, and believe that she will be able to help improve decision-making processes. One editor said:

> I guess from what I've seen, [her]"style", ability to seek consensus, make compromise and also get backing for some executive decisions seems solid. And [she] has been around for a long time, is obviously committed, and understands what's happening in development as well as style/editing. [30]

Editors also see her role as a necessary one, despite attempts to "democratize" the decision-making process: "People get easily distracted in the kind of discussions that are happening on the style mailing list so there has to be somebody to moderate the discussions and make decisions" [28].

Another area where disagreements must be resolved is in the editing and voting process itself. Edit notes allow established editors to help new editors understand the style guidelines by letting an editor know what he has done wrong or could do better. It is a place for user-to-user support [44], where editors help other editors because they want to ensure that things are done correctly, according to the standards set by the community, and that editors uphold the normative level of quality and accuracy.

But this helpfulness is not always interpreted as such, and not all editors come across as kind. Survey respondents were asked to rank their sentiment on the question, " I have had to defend my edits to other people" (Table 12). The mean ($\mu = 3.2217$) shows that, indeed, respondents largely agree with this statement and have had to defend edits to others.

Table 12: I have had to defend my edits to other people.

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
I have had to defend my edits to other people.	3.1127	45	137	29	4	15

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Despite the sometimes "terse"—as one editor characterized it—demeanor exhibited by some of the site's editors, editors continue to edit the data. One editor he believes that dealing with these conflicts is the price he pays for access to and participation in the cultural commons. "I wouldn't say if somebody was totally really a jackass toward me that I would quit MusicBrainz. It would be also highly unlikely that I would stop contributing" [36]. Conflict, then, is a tradeoff, a personal cost outweighed by the benefit of the cultural resource.

But participating in the background decision-making processes, like the Style Council, prove to be a different kind of challenge: "It's harder, and getting people to work on constitutional type

things, in terms of not just editing their favorite artists but coming up with rules for people to edit and, this is the really hard part, getting other people to agree, which is always the part that is hard." [39] Editors hope that the community manager will help make the process more inviting to editors by moderating conflict and focusing conversation.

D. Attitude & Motivation

The Importance of Open Source

One of the two categories of intrinsic motivation Lakhani and Wolf identify is "obligation/community-based intrinsic motivation," where "the goal to act consistently within the norms of a group can trigger a normative frame of action" [55:5]. One of these norms tested in Lakhani and Wolf's study is that "source code should be open." MusicBrainz editors were asked similar questions, both on the survey and in interviews. Interview and survey data indicate that MusicBrainz editors are motivated by obligation/community-based motivations, specifically the values of open source and Creative Commons. Editors feel that they are working to build a valuable information resource that they want to use freely and want others to have access to as well.

The MusicBrainz survey asked three questions aimed at measuring a similar belief to "source code should be open" among MusicBrainz contributors, but in relation openness of metadata. The results are shown in table (Table 13). Both registered and unregistered editors were asked to answer the first statement, "Information resources like MusicBrainz…should be free," so results have been filtered to include only the answers of registered editors.

Table 13: Open Source Likert-type Questions (Registered Users Only)

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
Information resources like MusicBrainz, peer-produced or otherwise, should be free.	3.7013	165	63	3	0	2
Metadata should be free and open to all	3.7076	172	59	5	0	3
Users who tag their files using MusicBrainz should contribute something to the database.	2.4978	19	89	99	19	9

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

The results to the first two questions shown in the table demonstrate that editors overwhelmingly believe that data should be free, consistent with open source principles as an obligation/community-based motivation [55]. However, respondents are split almost evenly (μ = 2.4978) on the statement, "Users who tag their files using MusicBrainz should contribute something to the database." About half of the respondents are content with anyone being able to take the data without giving back, or free-riding. The other half do not believe that free-riding is acceptable and would like to see everyone who uses the database give something back to it. It is also important to note that the taggers are open source-licensed as well, and some respondents

work on MusicBrainz-related software, but these three questions focus specifically on the use of the data.

Editors further elaborated on the importance of MusicBrainz as a cultural commons. One editor called it only "fair" that data contributed "by thousands of volunteers" should be free for those volunteers to use [31]. Another editor called the openness a "philosophy," and feels like contributions are part of building a bigger and better project that is useful to others. The same editor also noted the history of MusicBrainz as "a good example of what I'd be afraid of in another situation and would prevent me from contributing to another service" [37]. Another editor feels that "I get to benefit from MusicBrainz and this is somewhat of my payback to the community." [36]

One editor describes the process of using data and adding data as a "virtuous circle." He contributes to a variety of open source and peer-produced projects, including Wikipedia, and called himself "selfish" when asked why he contributes to open source projects. He said, "I think people who really value things will want to ensure they continue. And there are two ways you can do it. One, you can use your wallet. The other one is, if it's an option, you can contribute and make it a better thing" [33].

In the case of MusicBrainz, some editors do contribute with their wallets. A "Donate!" link appears in the left-hand sidebar of the non-NGS site, and a user can click to "Donation check" from their profile pages. Once a user has donated, the message displayed says, "You have donated and will not be nagged for another [#.#] days. Thank you for your support!" (Figure 10: Donation Check Message).

Figure 10: Donation Check Message

The MetaBrainz foundation recently held a fundraiser to raise $15,000 for new servers to support the upgrade to NGS. According to the MusicBrainz blog, individual users donated $3708.50, second only to the $5,000 donated by Google. Editors are not only willing to donate their time to the data, they are also willing to donate money in order to support the community's efforts and ensure that MusicBrainz continues to exist.

Sense of Community

A sense of belonging to "something bigger" motivates many editors and the community of editors fosters that sense of belonging. Editors cooperate both with editors of similar taste and editors with completely different taste to build the database. The community overall does not judge editors based on musical taste, and they understand the importance of people liking different kinds of music to the growth of the database. In addition to the community being open and accepting, several editors in the inner core have even developed friendships.

Because community-based motivations are at play, it is important to understand what the community looks like and how editors relate to one another. Borrowing an analogy from a friend, Robert Kaye describes the MusicBrainz community as an onion, from the outside in. Based on Kaye's description and interpretation of qualitative data, the layers are laid out below and named based on characteristics of contribution and community involvement. They are qualitatively ordered based on the amount of free time each level spends on the project:

— Layer 1: Free-riders — People who find MusicBrainz because they need to clean up their metadata, complete the task and never come back.
— Layer 2: Transients — People who make at least a few contributions.
— Layer 3: Fans — People who get really into the database, "adopt" their favorite artists, and make sure that the music by those artists is properly covered.
— Layer 4: Non-Technical Socialites — People who, from the fan stage, become involved with the MusicBrainz community and begin working on non-technical tasks, like documentation and style guidelines.
— Layer 5: Technical Socialites — People who write code and make a variety of more technical contributions.
— Layer 6: Inner Core — People who dedicate a lot of their free time on MusicBrainz, help other users, hang out in IRC or on forums and mailing lists, and "are likely to get hired into MusicBrainz."
— Layer 7: Employees — The founder, the two developers, and a community manager who are paid to work on MusicBrainz, plus the MetaBrainz Board of Directors. [49]

This core group of users, consisting of Layer 6 and Layer 7, is incredibly tight, and many of these editors can be found in the *#musicbrainz* IRC channel. When asked, "What makes the core people 'core'?" one editor shared his perspective:

> They are helping to build MusicBrainz, not just the MusicBrainz database. The community of editors is a little bit different. As you edit data in MusicBrainz, you will often notice the same people voting and commenting on your edits and you know you can ask these people about their opinion about some stuff. [28]

The core editors are heavily engaged in the community, and help other users across the site. An editor commented, "It's a very accepting community, I mean I've only been working with these people for five months or something, and they're saying, 'here go run this meeting for us for this thing that's important to our project that we've been working on for 2 years," which is really nice, to be able just to dive in" [37].

Editors in the Inner Core have also developed friendships. As one employee put sit, "I know the people I'm talking with all the time because I've obviously been around for about 3 years now and I've got quite a few friends" [35]. The same editor has attended several MusicBrainz meet-ups, hosted by Kaye, which give editors the opportunity to meet and socialize face-to-face. According to this editor, the first time he attended a meet-up, most of the conversation revolved around music and MusicBrainz. The second time around, the conversation had come to include

nearly anything.

Conversations in the IRC channel also extend beyond MusicBrainz. Editors in the channel talk about what they're eating, what they're listening to, or just pop in when they happen to be looking for some company. For example, on February 19, a Friday, several users engaged in a discussion about what they'd been drinking that evening. And on February 23, one editor talked with other editors in the channel about a German assignment she was working on, writing, "*[editor] hates writing about herself.*" Another editor engaged in a discussion with her about how he also hates writing about himself, and about a similar class assignment he had to endure in the past. Observing one month of chat logs also shows that there are very long periods of silence, where people are logged in but no discussion occurs.

The Likert-type question statement, "As a contributor, I feel part of a community and its mission" sought to get a sense of how closely tied editors believe they are to MusicBrainz as a community and not just a cultural commons. The results are shown in Table 14: As a contributor, I feel part of a community and its mission.

Table 14: As a contributor, I feel part of a community and its mission.

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
As a contributor, I feel part of a community and its mission.	3.1598	66	136	22	4	3

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Results from this question were run against two other variables in an attempt to show association between time and sense of community. First, the variable *login,* which contains responses to the categorical survey question "When did you last log in to MusicBrainz?" was split into two groups: Those who logged in within the last year and those who logged last year or more than a year ago. A t-test run against the community mission responses (variable: *communitymission*) did not reject the null hypothesis, showing no statistical difference between the means. Second, an ANOVA run with the four categories the variable *memlength* (discussed earlier) also failed to show any significant variance. None of the data collected through the survey was able to show a correlation between login frequency or membership length and feeling a sense of community. Editors of all membership length and login frequency are likely to feel a high sense of community. The survey did not measure the amount of free time spent on the project, which would potentially show that users who spend more free time on the project tend to feel a heightened sense of community. However, in the next section, a hypothesis relating to Auto-Editors provides insight into an answer (see **Auto-Editor Hypothesis D).**

As mentioned in the previous section, Patterns & Processes, users who share musical taste often come together to cooperate on a niche within the database. This cooperation fosters a special sense of community for the editors working together. One editor explained, "When I started, I knew a few editors working on the artists that I worked on so we collaborated, reviewing each others edits. It was a kind of nice experience" [28]. But even if an editor has not found a cohort with which to collaborate closely, the community will still provide help without judgment regarding taste: "If you ask for help with a particular metal album you'll never get nasty

comments from people who don't like metal or anything. They are all people who care about the data and want to help make the database better" [31].

As noted in the demographic section, this sample was predominately male, but one female editor participated in the interviews, and spoke briefly about what it is like to be a woman in a male-dominated community. She says that she has heard that guys are more direct than women in terms of communication and says, "I've never been a typical girl, but I get the feeling most girls would be a bit daunted by being surrounded by only guys." [29].

Reward & Responsibility

MusicBrainz does not have any sort of formal achievement or reward structure in place, but all editors in interviews expressed that they notice active users and begin to "trust" them at some point, demonstrating that editors build reputation through quality contribution [20]. Recognizing high-quality and frequent contributors, certain editors receive the title of Auto-Editor. An Auto-Editor's votes are instantly committed to the database and do not go through the voting process, unless an Auto-Editor chooses to turn off auto-editing. Auto-Editors make more contributions than non-Auto-Editors, are more inclined to help new users learn the ropes, and feel a greater sense of belonging to the community and its mission.

One editor told me, "I remember when I was starting out I'd see other people that were auto-editors thinking like wow, I'd love to get that one day, you're like respected or know what you're doing" [35]. In addition to "reasonably significant volume of current activity," Auto-Editors must have "a willingness to help and support other users" and demonstrate objectivity in doing so. Any current Auto-Editor may nominate another editor for the title by clicking "Nominate for auto-editor" on an editor's profile page. Two other Auto-Editors must then second the nomination, and the other Auto-Editors then have a week to vote in a general election.

The title of Auto-Editor is more like the collective identity of "hacker," an intrinsic, obligation/community-based motivation, than it is like an extrinsic motivation such as receiving pay or building professional skills [55]. It is a title earned by reputation, an important community-based motivation factor [20]. "Auto-Editor" is a "badge of honor," displayed on a user's profile [55]. While all paid employees have the title of Auto-Editor, the developers and the community manager earned the title from the community before being hired to work for pay.

Of the 214 respondents who answered the question, "Are you an Auto-Editor?" 70 responded affirmatively. During the period in which the survey was open, there were 191 total Auto-Editors, a number compiled from the Auto-Editors list page on April 5, 2011 [11], with those who were not Auto-Editors before July 24, 2010 (date the survey closed) subtracted (nine in total). Election dates are publicly available on the elections page [12]. Roughly 36% of editors who were Auto-Editors during the period of July 8 – July 24 completed the survey. The length of time it takes before nomination varies, but one Auto-Editor explained, "I was pretty active in it, I guess I started in like February of 2004 or something like that, I think within about 6 months I had done enough edits that I got the attention of other Auto-Editors and got nominated" [39].

The title of Auto-Editor is a reward for establishing a reputation of being a solid contributor to MusicBrainz. Sustaining this reputation then motivates editors to continue their contributions,

and to fulfill the responsibilities of the role that has been delegated to them by the other users of the site, including helping other users [44]. According to responses to "I want to be recognized by the MusicBrainz community for my contribution," receiving recognition for contribution is also more important to Auto-Editors than non-Auto-Editors (see Table 15).

Table 15: Likert-type Recognition (Auto-Editors vs. Non-Auto-Editors)

	R is an Auto-Editor		R is not an Auto-Editor			
	N	Mean (SD)	No	Mean (SD)	t	Cohen's d
I want to be recognized by the MusicBrainz community for my contributions.	63	2.84127 (0.5738)	126	2.5635 (0.7642)	2.5472+	0.4110

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree
diff = mean(yes(1)) – mean(no(2));; +SD inequality required Welch's unequal t-test
p(T>t) = 0.0029 (highly significant); Cohen's *d* indicates moderate effect size

Part of the responsibility of an Auto-Editor is to help new editors learn the guidelines and do things correctly. Thus, the title of Auto-Editor comes with an expectation that the editor will engage (and likely has already engaged) in supporting other editors [44]. It is a motivation to increase or to continue levels of support in order to acclimate new editors and teach them the norms of the project. This leads to the first two of several hypotheses relating to the motivations and feelings of Auto-Editors versus non-Auto-Editors.

Auto-Editor Hypothesis A: Due to the "willingness to help and support other users" requirement, Auto-Editors are more likely to engage in the activity of making notes on edits and votes than non-Auto-Editors. Results are shown in

Table 16: Chi-Square Table, Auto-Editor Hypothesis A

		R has made notes on edits and votes		
		Yes	*No*	*Total*
Auto-Editor?	*Yes*	68 (64 expected)	1 (5 expected)	69
	No	122 (126 expected)	14 (10 expected)	136
	Total	190	15	205

Pearson's Chi-square(1) = 5.285, *p* = 0.022; Cramer's V = 0.1605 (weak association)

A Pearson's Chi-Square test shows a *p*-value of 0.022, which rejects the null of independence and shows statistical significance, so there is some dependence between the two variables. However, the Cramer's V of 0.1605 shows only a weak association, so a larger sample of non-Auto-Editors might show greater association. The association in this sample is too weak to make any conclusive statements.

Auto-Editor Hypothesis B: Because "willingness to help and support other users" is a trait required of Auto-Editors, Auto-Editors should have a higher mean of agreement with the statement "I try to help new users understand the community rules and guidelines" than non-Auto-Editors. Results shown in Table 17.

Table 17: t-test Results, Auto-Editor Hypothesis B

	R is an Auto-Editor		R is not an Auto-Editor			
	N	Mean (SD)	N	Mean (SD)	*t*	Cohen's *d*
I try to help new users understand the community rules and guidelines.	63	3.285714 (0.5514)	101	2.940594 (0.7722)	3.3313	0.5143

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree
diff = mean(yes(1)) – mean(no(2)); +SD inequality required Welch's unequal *t*-test
p(T>t) = 0.0005, highly significant; Cohen's *d* indicates medium practical significance

A *t*-test comparing the mean of responses of Auto-Editors to non-Auto-Editors to the question "I try to help new users understand the community rules and guidelines" rejects the null hypothesis and favors the one-tailed Auto-Editor Hypothesis B. There is high significance with a *p*-value of 0.0005, and a Cohen's *d* of 0.5143 shows a medium effect size, indicating moderate practical significance.

Auto-Editors earn their title through continued quality contribution to the site. A large number of edits, either in a single area or across the site, draws the attention of other Auto-Editors and leads to a nomination. This leads to the next Auto-Editor hypothesis.

Auto-Editor Hypothesis C: Because of the high level of activity required for nomination, and because of the responsibility tied to the title, Auto-Editors should have a higher mean of edits entered than those who are not Auto-Editors. Additionally, because of the correlation between the number of edits an editor has entered and the number of artists to which an editor subscribes (see Patterns of Contribution), Auto-Editors should also have a higher mean of artist subscription than those who are not Auto-Editors.

Table 18: t-test Results, Auto-Editor Hypothesis C

	R is an Auto-Editor		R is not an Auto-Editor			
	N	Mean (SD)	N	Mean (SD)	*t*	Cohen's *d*
**Log of Total Edits Entered (logentered)	59	9.347487 (1.8483)	110	6.500057 (1.642767)	10.2773	1.6285
**Log of Total Artist Subscriptions (logsubs)	57	6.081008 (1.850873)	74	4.56101 (1.978707)	4.4823	0.7934

diff = mean(yes(1)) – mean(no(2)); **p(T>t) = 0.0000

Using two separate *t*-tests (results shown in Table 18) the means of the two groups were compared on the log of total edits entered and the log of artist subscriptions (log used because variables were not normally distributed). Both tests reject the null hypothesis and show and that the Auto-Editors indeed have a higher mean than non-Auto-Editors, with high significance (p = 0.0000). Cohen's *d* indicates a large effect size, especially in the case of the number of edit entered. These results support the hypothesis that Auto-Editors tend to make more edits and have more artist subscriptions than non-Auto-Editors, consistent with Auto-Editor as a reward for participation.

Finally, Auto-Editors must be noticed and nominated by the community, thus it is a kind of recognition for participation. This leads to Auto-Editor Hypothesis D.

Auto-Editor Hypothesis D: Because of the way contributes to the level of community involvement, and because Auto-Editors nee to spend a lot of time on the project and be recognized by the community, Auto-Editors will be more likely to relate positively to the statement, "As a contributor, I feel part of the community and its mission."

Table 19: t-test Results, Auto-Editor Hypothesis D

	R is an Auto-Editor		R is not an Auto-Editor			
	N	Mean (SD)	No	Mean (SD)	t	Cohen's d
As a contributor, I feel part of a community and its mission.	68	3.3235 (0.5309)	137	3.07299 (0.7137)	2.8255	0.398337

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree
diff = mean(yes) – mean(no); +SD inequality required Welch's unequal *t*-test
p(T>t) = 0.0026; Cohen's *d* indicates medium effect size

A Welch's unequal *t*-test (Table 19) rejects the null hypothesis and favors the one-tailed Auto-Editor Hypothesis D, and with a *p*-value of 0.0026, it is highly significant. The Cohen's *d* of 0.398337 shows only a medium effect size, but still indicated practical significance. Thus, Auto-Editors are more likely to feel a sense of community than non-Auto-Editors. This may also satisfy the community questions in the Sense of Community section of this paper, as Auto-Editors spend an extraordinary amount of free time working on the project.

The Auto-Editor title is a reward for contribution, but it also carries a heavy burden of responsibility. Where normal editors must see their edits through the voting process, edits made by Auto-Editors are—as the title indicates—immediately accepted in the database. One user explains:

> I didn't exactly turn it down but I was a bit unsure about taking it because it does feel like a title that you've earned, but it also feels like quite a bit of responsibility or at least it did to me when I got it. And I wasn't entirely sure if I was ready to have an auto-editor title. [35]

The Auto-Editor page displays the quote, "With great power comes great responsibility," credited to "Uncle Ben (of Spider Man fame)" [13], using a pop culture reference to ground the importance of the title. The same page also notes that while expertise in a specific genre, style, label, or other area of music is good to have, Auto-Editors should also be able to "edit outside their realm of expertise." This indicates that Auto-Editors must contribute to both the breadth and the depth of the database.

Barriers and Tradeoffs

MusicBrainz has a steep learning curve for new editors, both technically and in terms of norms. But for most editors, any barriers to entry are outweighed by the fact that they are able to edit the data directly. Being able to correct data or add missing data makes editors feel better about the quality, and satisfying both their information need—extrinsic motivation—and their sense of obligation to the mission of the constructed cultural commons.

Because of the great detail to which editors have gone to negotiate fields and metadata structure and representation, the style guidelines are quite complicated—and daunting. The best way to learn the guidelines is by editing and taking the feedback of other editors. One editor said, "I got a few things wrong and [an editor] pointed me to the relative guideline, et cetera, I read it, I got it, I got why it was there, I understood, I applied it, and then I continued" [27].

The interface can be a challenge in and of itself. "It's a bit clunky, the interface," explained one editor. "You have to jump through quite a lot of hoops to edit existing information or even to add new information" [36]. The following exchange from the IRC room is a telling example of the difficulties new editors have with the user experience:

```
01:59:11 <mchou_> Failbus
01:59:16 <mchou_> no can do
01:59:22 <mchou_> I'm giving up
01:59:24 <nikki> :/
01:59:34 <mchou_> "Select more release groups to merge."
01:59:37 <mchou_> lol
02:00:04 <mchou_> It doesn't tell me how to do that
02:01:19 <mchou_> freaking lame
02:02:00 <nikki> if you give me links, I could merge them
02:02:18 <nikki> * nikki isn't sure how to explain it in a way that
works for you :/
02:02:45 <mchou_> http://musicbrainz.org/release-group/ae3c0601-5838-
4fff-9ba2-946514e1f26d.html
02:02:55 <mchou_> http://musicbrainz.org/release-group/9a8f71d4-a551-
47fb-9cbd-bece4ed098eb.html
02:03:15 <mchou_> after you do it explain to me how it's done from
those links
02:03:55 <mchou_> cause I think the inteface is brain dead
```
(Chat Log, 2/25/11)

However, most editors in interviews found the interface manageable. One editor called it "decent" and during our interview showed me—a new editor—how to use the track parser to more easily enter new releases, yet another example of editor-to-editor support. Another editor found the style guidelines far more challenging to master, saying, "I had no trouble with the user interface itself. It took a little longer to become familiar with all the guidelines a user is supposed to adhere to" [31].

But users see this learning curve as "the price you have to pay in order to use it." The same editor called MusicBrainz "the only accurate source of information" [36]. All editors emphasized the importance of data quality in interviews, often in comparison to proprietary sources. One editor commented, "a lot of times when there are mistakes in CDDB or something there is no

way to correct them. There's no way to fix it or override it versus MusicBrainz, you can suggest edits and people vote on it. It's much more open" [26]. Another user rejected both of the major proprietary sources, calling their data "incomplete" and "crap" [30]. A few MusicBrainz editors commented on inconsistencies in Amazon's commercial data, like misplaced parentheses in track titles for re-releases. The global nature of the editors also means that the coverage spans far beyond the American music industry, covering genres, artists, and styles around the world.

With MusicBrainz, people are free to make any necessary corrections as long as they create a MusicBrainz account. The gratification is nearly instant, and although not committed to the database instantly, changes are immediately visible on an editor's page. The hands-on ability to change things is a common theme among editors who emphasize the importance of high data quality (see Table 20, first question, showing that editing the data helps most users feel better about the quality).

Table 20: Likert-type Questions relating to Accuracy and Quality of Contributor

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
Editing MusicBrainz helps me feel better about the accuracy of the data.	3.3671	93	138	6	0	1
Anyone is qualified to add & edit data about music in MusicBrainz.	3.1579	37	140	54	7	2

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Another barrier to contribution relates to musical taste. A new editor who listens only to popular music is less likely to find work to do in the system than one who listens to music that is more obscure. As a participant-observer in the study, I found a niche in the independent music released by a few labels that send me promotional downloads of albums. Because I get the music before others do, and because many of the bands are new or relatively unknown, I am able to add these releases before other editors do.

Despite the emphasis on high quality, the coverage of popular music, and the difficulty learning the interface and the guidelines, however, editors believe that anyone is qualified to participate edit data about music in MusicBrainz (see Table 20 where the mean shows that editors feel positively about anyone editing data). The checks and balances put in place by the community are such that new editors receive constant support from experienced editors. Plus, editors who may not be able to contribute to the breadth of the database with new releases can always contribute to depth by finding typos, misspellings, and relationships to be added and corrected.

The Act of Editing

Lakhani and Wolf explain that "having fun or enjoying oneself when taking part in the activity is at the core of the idea of intrinsic motivation" [55:4]. They discuss the work of psychologist Csikszentmihalyi, who wrote about "flow" in activities, part of which means that "enjoyment is maximized, characterized by intense and focused concentration" [55:4]. Enjoyment-based

intrinsic motivation was one of the primary motivations for F/OSS contributors in their study. Similarly, enjoyment-based motivation drives MusicBrainz editors to contribute to the commons. Editors also demonstrate a compulsion to fix bad data, and the phrase "obsessive-compulsive disorder" is a colloquialism used to explain what makes editors edit.

Editors in interviews described the act of editing data as "fun," "relaxing," and "meditating." For one user, "reading about music, knowing that I'm helping to build 'something big'" [28] crosses over into obligation/community-based intrinsic motivation as well. The results of two survey questions shown in Table 21 illustrate that while editors moderately agree that MusicBrainz is "intellectually stimulating," a strong majority believe that contributing to MusicBrainz is "fun," supporting the claim that enjoyment-based motivation is at play here.

Table 21: Likert-type questions relating to the act of editing

	Mean	Strongly Agree	Agree	Disagree	Strongly Disagree	No Answer
Contributing to MusicBrainz is an intellectually stimulating activity	3.0046	55	122	45	7	5
Contributing to MusicBrainz is fun	3.3247	102	108	15	6	3

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree

Obsessive-compulsive disorder is a colloquialism within the MusicBrainz community, and many of the interview subjects referred to themselves and other editors as "OCD," "obsessive compulsive," or expressed similar sentiments:

- "I started editing slowly and kind of got hooked really I guess and sort of OCD, I had to get all of my music on there" [35]
- "I'm kind of a little OCD about it, I want it to be right." [26]
- "Well the Auto-Editor is basically like a badge of obsessive compulsiveness" [39]
- "I started editing even albums that I didn't actually have, just to clean up data on MB, which turned out to be very addictive" [28]
- A lot of places I think don't have any verification system and aren't open to editing by anyone who's willing to be OCD enough to correct the data, which is nice… the more compulsive you're willing to be about things like scanning in all of your discs and transcribing the text that's in that little circular thing around it could actually end up being useful for a database like MusicBrainz." [37]
- "I don't know, it just doesn't look right if you have two discs from an album and one of them uses parentheses and the other brackets, e.g. (disc 1) and [disc 2]" [31].
- "I became kind of obsessive for a while… if I came to an artist page, I'd see things that were obviously wrong or against the style guidelines and things and then I'd start cleaning it and then it would lead to, "Oh, it doesn't have album art, I need add that" and "Oh, it doesn't have the release dates, where can I get the release dates from?" and "Oh, it's not linked to Wikipedia, not linked to Discogs, not linked to everything else." And suddenly if it wasn't complete it was frustrating to me and I had to add everything I could [30].

These users are not clinically diagnosed as OCD —"Do I need to wash my hands after? No" [32]—but they use it as a way to explain their willingness to spend time agonizing over details that those outside the community would see as meaningless. They value neatness and order for practical reasons, but shrug it off as a personality trait that the editors share in common. One editor notes, "Sometimes when it's just I have some time to kill with nothing else to do I just try to somehow get it in order" [36].

The OCD phenomenon as motivation is both extrinsic and intrinsic in that editors feel *compelled* to correct the data—a kind of compulsion-based motivation—and extrinsic in that this extreme attention to detail carries over to their own music collections [55]. Many editors described very regimented processes by which they add new music to their collection, and getting the data right in MusicBrainz is a part of that tagging process. It makes it a better information resource, thus satisfying their information need.

E.   Future Work

This paper is an all-encompassing summary of the background, methodology, and research findings from a one-year study of the MusicBrainz editor community. It is a case study of a cultural commons, contributing to the growing pool of research on peer-produced, commons-based information resources. Understanding how and why people contribute to cultural commons, how and why the commons are able to persist, and how the commons operate can inform future cultural commons, an important development enabled by digital communication technologies.

The vast amount of data collected and the breadth of this paper make possible a number of follow-up papers and focused areas of continued study:

1. Comparative Analysis of Wikipedia and MusicBrainz: This case study could be compared to one of any number of case studies conducted on Wikipedia, either broadly structured to cover a swath of motivation or narrowly focused on specific patterns of activity.

2. Comparative Analysis of F/OSS Projects and MusicBrainz: This case study could be compared to one of any number of case studies conducted on F/OSS software projects or open source technical projects. Like a Wikipedia comparison, the analysis could be structured broadly or confined to narrow areas of focus.

3. Comparative Analysis of MusicBrainz and other open source music information sites: As discussed in this paper, there exist other cultural commons for music metadata. What are the characteristics of these commons? Are they narrowly focused on specific genres or types of music? Do the communities consist mainly of consumers or creators of music? How do their structures compare to MusicBrainz?

4. Women and Commons-based Peer Production: The disparity between genders in my sample is itself an area of further inquiry. Why don't more women participate? Is this true

of all cultural commons or just some commons sharing certain characteristics? What are those characteristics?

5. Metadata Structure and Representation: As we move closer to the establishment of the celestial jukebox—for example, with Amazon's cloud service for music—metadata becomes an increasingly important piece of the MIR puzzle. In what ways could the deeply contextual and highly structured data in MusicBrainz become ingrained in the future of MIR? How do other sources compare in terms of their structure and representation? A survey of metadata sources, or simply a look at the semantic capabilities of MusicBrainz, would be a rich area of further inquiry.

6. Cultural Commons Frameworks: Madison et. al.'s framework for studying constructed cultural commons could be applied to this MusicBrainz case study, making it easier to compare the results of this study to the results of other studies of cultural commons. This paper is currently in progress.

7. Mapping Last.fm usage to MusicBrainz participation: Due to the link between MusicBrainz editors and Last.fm users, there is potential to look more deeply into how the two play off of each other. How do users' statements about their musical taste compare to what they display on Last.fm? How do MusicBrainz editors manage their Last.fm profiles? What characterizes the relationship between the two sites?

V. Conclusion

The shift to the Next Generation Schema comes with a rebrand of the project. Where the current site reads "community music metadatabase," the NGS Release Candidate test site touts MusicBrainz as "The Open Music Encyclopedia." It also contains new descriptions of MusicBrainz' mission:

> 1. **The ultimate source of music information** by allowing anyone to contribute and releasing the data under open licenses.
>
> 2. **The universal lingua franca for music** by providing a reliable and unambiguous form of music identification, enabling both people and machines to have meaningful conversations about music. [2]

The second point, "enabling both people and machines to have meaningful conversations about music" points to a potential push toward the Semantic Web capabilities of MusicBrainz [74], leveraging data structures like XML and RDFa, discussed earlier in this paper. One editor remarked that he stopped thinking about MusicBrainz as a tagging solution and started to think about it as a Semantic Music Encyclopedia: "Something like a structured Wikipedia—that is, something that can be processed by machines and helps building various tools on top of it" [28].

MusicBrainz will continue to grow as long as it can find contributors driven by the necessary motivations—both intrinsic and extrinsic—to overcome social dilemma equilibrium and cooperate. MusicBrainz editors are a hybrid of F/OSS contributors and Wikipedia contributors. They are driven to contribute to the Wikipedia-like pool of cultural knowledge by intrinsic motivations like enjoyment, community obligation, and the ability to share musical knowledge; but, like F/OSS projects, MusicBrainz satisfies an information need, and adding knowledge to the database makes it a more useful tool for editors themselves as well by storing metadata about music they own.

Editors commented on the notion of "completeness," and all agree that, due to the nature of music as an ever-evolving cultural art form, MusicBrainz "complete" is unattainable. Because new music is constantly created, MusicBrainz will never be complete, regardless of how many editors contribute to the commons. This will not, however, dissuade editors from striving to achieve the most accurate, complete, and high-quality source for music metadata available, especially as technologists only begin to leverage its semantic capabilities. As a peer-produced information resource for music metadata, MusicBrainz is an important model of a thriving cultural commons.

VI. Acknowledgements

VII. References

1. Welcome to MusicBrainz! *MusicBrainz*. http://musicbrainz.org/.
2. Release Candidate Test Page. *MusicBrainz*. http://test.musicbrainz.org/.
3. AboutMusicBrainz. *MusicBrainz*. http://musicbrainz.org/doc/AboutMusicBrainz.
4. Social Contract. *MusicBrainz*. http://musicbrainz.org/doc/Social_Contract.
5. AllMusic. http://allmusic.com/.
6. All genres. *SoundUnwound*. http://www.soundunwound.com/genre/cloud?ref=F.
7. LiveDataFeed. *MusicBrainz*. http://musicbrainz.org/doc/LiveDataFeed.
8. HowPUIDsWork. *MusicBrainz*.
    http://musicbrainz.org/doc/HowPUIDsWork#How_Picard_uses_PUIDs.
9. Beginners Guide. *MusicBrainz*. http://test.musicbrainz.org/doc/Beginners_Guide.
10. Proposal:Revised Sortname Style. *MusicBrainz Wiki*.
    http://wiki.musicbrainz.org/Proposal:Revised_Sortname_Style.
11. The MusicBrainz AutoEditors. *MusicBrainz*. http://musicbrainz.org/mod/automodlist.html.
12. Auto-Editor Election. *MusicBrainz*. http://musicbrainz.org/doc/Auto-Editor_Election.
13. AutoEdiing metor. *MusicBrainz*. http://musicbrainz.org/doc/AutoEditor.
14. Anthony, D., Smith, S., and Williamson, T. Explaining Quality in Internet Collective Goods:
    Zealots and Good Samaritans in the Case of Wikipedia.
    http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf.
15. Axelrod, R. *The Evolution of Cooperation*. Basic Books, New York, 2006.
16. Bernard, H. *Social Research Methods: Qualitative and Quantitative Approaches H. Russell
    Bernard*. SAGE, Thousand Oaks Calif. [u.a.], 2005.
17. Bimber, B., Flanagin, A.J., and Stohl, C. Reconceptualizing Collective Action in the
    Contemporary Media Environment. *Communication Theory 15*, 4 (2005), 365-388.
18. Cano, P., Batlle, E., Gomez, E., de C.T.Gomes, L., and Bonnet, M. Audio Fingerprinting:
    Concepts and Applications. In *Computational Intelligence for Modelling and Prediction*.
    Springer-Verlag Berlin/Heidelberg, [New York], 2005.
19. Charmaz, K. *A Practical Guide through Qualitative Analysis*. SAGE, London, 2006.
20. Cheshire, C. and Antin, J. The Social Psychological Effects of Feedback on the Production
    of Internet Information Pools. *Journal of Computer-Mediated Communication 13*, 3 (2008),
    705-727.
21. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates,
    Hillsdale N.J., 1988.
22. Creswell, J. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*.
    Sage Publications, Thousands Oaks, 2002.
23. Cunningham, S., Jones, M., and Jones, S. Organizing Digital Music for Use: An Examination
    of Personal Music Collections. (2004).
24. Dublin Core Metadata Initiative. DCMI Metadata Basics. *Dublin Core Metadata Initiative*.
    http://dublincore.org/metadata-basics/.
25. EchoNest & LabROSA. FAQ. *Million Song Dataset*.
    http://labrosa.ee.columbia.edu/millionsong/faq.
26. Editor1. MusicBrainz User Interview. 2010.
27. Editor10. MusicBrainz User Interview. 2011.
28. Editor11. MusicBrainz User Interview. 2011.

29. Editor12. MusicBrainz User Interview. 2011.
30. Editor13. MusicBrainz User Interview. 2011.
31. Editor14. MusicBrainz User Interview. 2011.
32. Editor2. MusicBrainz User Interview. 2010.
33. Editor3. MusicBrainz User Interview. 2010.
34. Editor4. MusicBrainz User Interview. 2010.
35. Editor5. MusicBrainz User Interview. 2011.
36. Editor6. MusicBrainz User Interview. 2011.
37. Editor7. MusicBrainz User Interview. 2011.
38. Editor8. MusicBrainz User Interview. 2011.
39. Editor9. MusicBrainz User Interview. 2011.
40. Feller, J. *Perspectives on Free and Open Source Software*. MIT, Cambridge, Mass., 2007.
41. Gordon, W. Response — Discipline and nourish: On constructing commons. *Cornell Law Review 95*, (2010), 733-756.
42. Gracenote. Open Source. *Gracenote: Developer Zone*. https://doors.gracenote.com/developer/open.html.
43. Gruber, T. Collective knowledge systems: Where the Social Web meets the Semantic Web✲. *Web Semantics: Science, Services and Agents on the World Wide Web*, (2007).
44. von Hippel, E.A. and Lakhani, K. How Open Source Software Works: 'Free' User-to-User Assistance? *SSRN Electronic Journal*, (2001).
45. Hsu, S.-L. What Is a Tragedy of the Commons? Overfishing and the Campaign Spending Problem. *SSRN Electronic Journal*, (2005).
46. International Federation of Library Associations and Institutions. Digital Libraries: Metadata Resources. *IFLANET*. http://archive.ifla.org/II/metadata.htm.
47. International Federation of Library Associations and Institutions. Functional Requirements for Bibliographic Records. *IFLA Cataloguing Section*. http://archive.ifla.org/VII/s13/frbr/frbr_current3.htm.
48. Jones, R. Fingerprinting and Metadata Progress Report. *Last.fm: The Blog*, 2008. http://blog.last.fm/2008/03/25/fingerprinting-and-metadata-progress-report.
49. Kaye, R. MusicBrainz Expert Interview. 2010.
50. Kaye, R. Please take the MusicBrainz Survey! 2010. http://blog.musicbrainz.org/?p=595.
51. Knees, P., Pohle, T., Schedl, M., and Widmer, G. A music search engine built upon audio-based and web-based similarity measures. *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*, (2007), 447.
52. Korfiatis, N.T., Poulos, M., and Bokos, G. Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review 30*, 3 (2006), 252-262.
53. Kranich, N. *Information Commons: A Public Policy Report*. NYU School of Law, Brennan Center for Justice, 2004.
54. Kvale, S. *Interviews: An Introduction to Qualitative Research Interviewing*. Sage, London, 1994.
55. Lakhani, K. and Wolf, R. "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. In J. Feller, B. Fitzgerald, S. Hissam and K.R. Lakhani, eds., *Perspectives on Free and Open Source Software*. MIT Press, Cambridge Mass., 2005.

56. Madison, M., Frishmann, B., and Strandburg, K. Constructing Commons in the Cultural Environment. *Cornell Law Review 95*, (2010), 657.

57. Mankiw, N. *Principles of Economics*. Thomson/South-Western, Mason Ohio, 2004.

58. Mark, N. Birds of a Feather Sing Together. *Social Forces 77*, 2 (1998), 453-485.

59. Marshall, C.C. and Shipman, F.M. Which semantic web? *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia - HYPERTEXT '03*, (2003), 57.

60. Mead, M. Music Is a Human Need. *Music Educators Journal 59*, 2 (1972), 24-29.

61. Mueller, J.H. The Social Nature of Musical Taste. *Journal of Research in Music Education 4*, 2 (1956), 113-122.

62. National Information Standards Organization. *Understanding Metadata*. NISO,, Bethesda, MD :, 2004.

63. Nov, O. and Kuk, G. Open source content contributors' response to free-riding: The effect of personality and context. *Computers in Human Behavior 24*, 6 (2008), 2848-2861.

64. Olson, M. *The logic of collective action : public goods and the theory of groups*. Harvard University Press, Cambridge Mass., 1965.

65. Ostrom, E. Analyzing collective action. *Agricultural Economics 41*, (2010), 155-166.

66. Prasarnphanich, P. and Wagner, C. Creating critical mass in collaboration systems: Insights from Wikipedia. *2008 2nd IEEE International Conference on Digital Ecosystems and Technologies*, (2008), 126-130.

67. Quach, L. and Trenz, D. MusicBrainz Expert Interview. 2011.

68. Radano, R. and Bohlman, Philip. Introduction: Music and Race, Their Past, Their Presence. In *Music and the Racial Imagination*. University of Chicago Press, Chicago; London, 2000.

69. Rafaeli, S. and Ariel, Y. Online motivational factors: Incentives for participation and contribution in Wikipedia. In A. Barak, ed., *Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia*. Cambridge University Press, Cambridge; New York, 2008.

70. Simson, J. The Cost of Bad Data. *Billboard*, 2010. http://www.billboard.biz/bbbiz/content_display/magazine/opinion/e3i6baa5818ddcc27e1167 17a5001718a1d?src=achallenge.

71. Smiraglia, R.P. Musical Works and Information Retrieval. *Notes 58*, 4 (2002), 747-764.

72. Suchman, L. and Jordan, B. Interactional Troubles in Face-to-Face Survey Interviews. *Journal of the American Statistical Association 85*, 409 (1990), 232-241.

73. Svenonius, E. *The Intellectual Foundation of Information Organization*. MIT Press, Cambridge Mass., 2000.

74. Swartz, A. MusicBrainz: a semantic Web service. *IEEE Intelligent Systems 17*, 1 (2002), 76-77.

75. TuneUp. TuneUp. *TuneUp*. http://www.tuneupmedia.com/.

76. Vellucci, S.L. Music Metadata and Authority Control in an International Context. *Notes 57*, 3 (2001), 541-554.

77. Yang, H.-L. and Lai, C.-Y. Motivations of Wikipedia content contributors. *Computers in Human Behavior 26*, 6 (2010), 1377-1383.

VIII. Appendix

The Appendix includes:
- The full MusicBrainz survey;
- Sort name discussion from the Style mailing list; and
- Chat log from February 23, 2011.

MusicBrainz: Community and Participation

As researchers interested in the motivations and communities powering open source projects, we would like to learn more about how and why people use MusicBrainz.
Welcome to the MusicBrainz Community and Participation Survey!

You work hard on MusicBrainz and we want to know why! As researchers at University of California—Berkeley's School of Information, this survey will help us learn more about the MusicBrainz community and what it takes to build and sustain a successful peer-produced database.

Your participation can help make MusicBrainz better!

The survey should take no more than 15 minutes to complete. We are grateful for your time, thank you!

There are 25 questions in this survey

Qualification

To take this survey you must be 18 years old.

---

1 In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old? *

Please choose **only one** of the following:

○ Yes

○ No

MusicBrainz Usage

These questions are about how and how often you use MusicBrainz.

2 Please indicate how often you engage in the following activities.

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Often	Fairly Often	Occasionally	Rarely	Only Once	Never
Use a MusicBrainz application to tag my data (e.g., Jaikoz, Magic MP3, Picard)	○	○	○	○	○	○
Use MusicBrainz as a resource to manually tag data	○	○	○	○	○	○
Search for release, artist, or label information on MusicBrainz	○	○	○	○	○	○
Contribute code or manage/track bugs in the MusicBrainz database	○	○	○	○	○	○

3 Have you ever discovered a new artist through the MusicBrainz database?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ Yes
○ No

4 Briefly, how did you make this discovery? Please describe.

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '3 [DISCOVERY 1]' (Have you ever discovered a new artist through the MusicBrainz database?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you

be at least 18 years of age. Are you at least 18 years old?)

Please write your answer here:

5 How did you learn about MusicBrainz?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

- ○ Friend
- ○ Blog or Twitter
- ○ Online discussion group or forum
- ○ Email list
- ○ Online ad
- ○ Creative Commons
- ○ Web search to clean up my music metadata
- ○ Other [   ]

6 How much do you agree or disagree with the following statements?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
MusicBrainz is a valuable resource for information	○	○	○	○	○

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
about music.					
MusicBrainz is a more complete resource than other databases available free online.	○	○	○	○	○
Information resources like MusicBrainz, peer-produced or otherwise, should be free.	○	○	○	○	○
MusicBrainz is valuable for discovering new music.	○	○	○	○	○
Peer-produced sources for music information are as reliable as non-peer-produced resources.	○	○	○	○	○
I am passionate about music.	○	○	○	○	○
Accurate and complete metadata enhances my music collection.	○	○	○	○	○

7 What media player(s) do you use to play music on your computer (e.g. iTunes, Songbird, WinAmp)?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please write your answer here:

Please type the names of any you use.

8 What websites do you use for information about music?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **all** that apply:

- ☐ Wikipedia
- ☐ Discogs
- ☐ All Music Guide
- ☐ Lyrics websites (e.g., LyricWiki)
- ☐ Online music publications (e.g., Pitchfork)
- ☐ Newspapers (e.g., New York Times)
- ☐ Magazines (e.g., Rolling Stone, Mojo, XLR8R)
- ☐ Radio stations and radio services (e.g., KEXP.org, Pandora, Last.fm)
- ☐ Podcasts (e.g., Sound Opinions)
- ☐ MP3 or music blogs (e.g., Fluxblog)
- ☐ Record store sites (e.g., Amoeba)
- ☐ Online music stores and services (e.g., Amazon, eMusic, iTunes)
- ☐ MusicBrainz
- ☐ Other: _____

9 Have you registered a MusicBrainz account? *

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

- ◯ Yes
- ◯ No

Activities

10 Which of the following activities have you performed?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Yes	Uncertain	No
Relating entities to non MusicBrainz URLs (Amazon, Wikipedia, etc.)	○	○	○
Editing information relating to MusicBrainz entities on Wikipedia, LyricWiki, or discogs	○	○	○
Adding releases to MusicBrainz	○	○	○
Working on a smaller community-defined project (e.g., Cleanup of the Month)	○	○	○
Editing documentation or informational pages on the MusicBrainz wiki	○	○	○
Participate in forum discussions or mailing lists	○	○	○
Adding/tracking/editing bugs	○	○	○
Contribute code	○	○	○

11 Which of the following activities have you performed?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Yes	Uncertain	No
Maintain a MusicBrainz server	○	○	○
Subsribe to new artists	○	○	○
Subscribe to editors	○	○	○
Vote on changes made by your editor subscriptions	○	○	○
Vote on changes made to your artist subscriptions	○	○	○
Vote on edits on the general edit list	○	○	○
Make notes on edits and votes	○	○	○
Add annotations	○	○	○

Attitude

12 How much do you agree or disagree with the following statements?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
Users who tag their files using MusicBrainz should contribute something to the database.	○	○	○	○	○
Anyone is qualified to add and edit data about music in MusicBrainz.	○	○	○	○	○
MusicBrainz has helped me expand my knowledge about music.	○	○	○	○	○
Metadata should be free and open to all.	○	○	○	○	○
Participating in MusicBrainz allows me to share my music knowledge with others.	○	○	○	○	○
Contributing to MusicBrainz is fun.	○	○	○	○	○

Resources

13 How much do you agree or disagree with the following statements regarding new users and MusicBrainz resources?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
It's important to provide guidance to editors in the notes box.	○	○	○	○	○
I try to help new users understand the community rules and guidelines.	○	○	○	○	○
The forums and site documentation provide helpful guidance.	○	○	○	○	○
As a contributor, I feel part of a community and its mission.	○	○	○	○	○
I want to be recognized by the MusicBrainz community for my contributions.	○	○	○	○	○
The community makes decisions without a lot of top-down management.	○	○	○	○	○
Editors who don't follow style and user guidelines frustrate me.	○	○	○	○	○

Contribution

14 How much do you agree or disagree with the following statements regarding your contributions to MusicBrainz?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
I feel a sense of accomplishment when I add data that wasn't already there.	○	○	○	○	○
Entities I edit reflect my taste in music.	○	○	○	○	○
"Yes" votes are the community's praise for good work.	○	○	○	○	○
Editing MusicBrainz helps me feel better about the accuracy of the data.	○	○	○	○	○
Contributing to MusicBrainz is an intellectually stimulating activity.	○	○	○	○	○

Editing

15 How much do you agree or disagree with the following statements regarding the editing process in MusicBrainz?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose the appropriate response for each item:

	Strongly Agree	Agree	Disagree	Strongly Disagree	Not Applicable
I prefer to focus on a few artists rather than browse all edits.	○	○	○	○	○
An editor should only add a release if s/he owns a physical or a digital copy.	○	○	○	○	○
I have had to defend my edits to other MusicBrainz editors.	○	○	○	○	○
I vote on edits only when I feel knowledgeable about the entity being edited.	○	○	○	○	○

User Information

A few questions how you use the site.

16 OPTIONAL: Please enter your username.

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please write your answer here:

We will only use your name to aggregate information about edits, votes, annotations, notes, and other contributions that are public to other logged-in MusicBrainz users. Our goal is to observe patterns of participation and interaction between users, and you will never be identified publicly in any way.

17 When did you register your MusicBrainz account?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please enter a date:

Your sign-up date appears next to "Member Since" on your profile page when logged in.

18 Are you an Auto-Editor?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ Yes

○ No

19 Are you paid to contribute to MusicBrainz?

Please choose **only one** of the following:

○ Yes

○ No

---

20 Approximately when did you last log into MusicBrainz?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ Today

○ Yesterday

○ Last week

○ Last month

○ Last year

○ More than a year ago

---

21 If you've ever met up with any other MusicBrainz users in real life, what were the circumstances?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **all** that apply:

☐ One-on-one

☐ MusicBrainz Summit

☐ Pub Meeting

☐ Other: [   ]

---

22 Please provide your email address if you would be willing to participate in a follow-up interview.

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '9 [REGISTERED BRANCH]' (Have you registered a MusicBrainz account?) *and* Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please write your answer here:

Your email address will remain confidential and be used only for us to contact you to arrange an interview.

Basic Information

These questions will help us understand demographic information about the MusicBrainz community.

23 How old are you?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ 18 to 21
○ 22 to 34
○ 35 to 44
○ 45 to 54
○ 55 to 64
○ 65 or Older

24 What is your gender?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ Female
○ Male

25 Where do you live?

Only answer this question if the following conditions are met:
° Answer was 'Yes' at question '1 [Qualify]' (In order to participate in this survey, we require that you be at least 18 years of age. Are you at least 18 years old?)

Please choose **only one** of the following:

○ North America
○ Europe
○ Asia
○ Africa
○ South America

○ Australia or New Zealand

○ Antarctica

Thank you for competing the MusicBrainz survey!

If you have any questions or comments regarding this survey, please contact Jess Hemerly at jhemerly [at] ischool [dot] berkeley [dot] edu.

24.07.2010 – 00:00

Submit your survey.
Thank you for completing this survey.

MusicBrainz-style Digest, Vol 72, Issue 1

1 message

---

musicbrainz-style-request@lists.musicbrainz.org <musicbrainz-style-request@lists.musicbrainz.org>
Reply-To: musicbrainz-style@lists.musicbrainz.org
To: musicbrainz-style@lists.musicbrainz.org

Fri, Apr 1, 2011 at 5:00 AM

Send MusicBrainz-style mailing list submissions to
    musicbrainz-style@lists.musicbrainz.org

To subscribe or unsubscribe via the World Wide Web, visit
    http://lists.musicbrainz.org/mailman/listinfo/musicbrainz-style
or, via email, send a message with subject or body 'help' to
    musicbrainz-style-request@lists.musicbrainz.org

You can reach the person managing the list at
    musicbrainz-style-owner@lists.musicbrainz.org

When replying, please edit your Subject line so it is more specific
than "Re: Contents of MusicBrainz-style digest..."

Today's Topics:

   1. Re: RFC 320: Revised Sortname Style (caller#6)
   2. Re: RFC 320: Revised Sortname Style (rossetyler)
   3. Re: RFC 320: Revised Sortname Style (rossetyler)

Message: 1
Date: Thu, 31 Mar 2011 15:34:13 -0700
From: caller#6 <meatbyproduct-musicbrainz@yahoo.com>
Subject: Re: [mb-style] RFC 320: Revised Sortname Style
To: musicbrainz-style@lists.musicbrainz.org
Message-ID: <4D950165.3060507@yahoo.com>
Content-Type: text/plain; charset=ISO-8859-1; format=flowed

On 03/31/2011 02:45 PM, Nikki wrote:
> caller#6 wrote:
>
>> Looking at the diffs, I see that prior to March 2010 the
>> guideline/example agreed with you, while the MB entry itself lists

>> "Hendrix, Jimi, The, Experience". I'm open to suggestions.
> Hmm... I think using your page, you could actually say:
>
> Family name: Hendrix
> Given name: Jimi
> Band name: Experience
> Leading article: The
>
> and therefore "Hendrix, Jimi, Experience, The".
>
> Nikki

Well, I had to at least /try/ to find the rationale for the existing
official example (which isn't spelled out).

Personally, I agree with you (and Paul), and will happily change the
example if nobody else can produce a compelling argument against doing so.

Alex / caller#6

Message: 2
Date: Thu, 31 Mar 2011 19:23:13 -0700 (PDT)
From: rossetyler <rossetyler@gmail.com>
Subject: Re: [mb-style] RFC 320: Revised Sortname Style
To: musicbrainz-style@lists.musicbrainz.org
Message-ID: <1301624593328-6229502.post@n2.nabble.com>
Content-Type: text/plain; charset=us-ascii

See related discussion here:
http://forums.musicbrainz.org/viewtopic.php?pid=12037#p12037

I think much more thought needs to go toward "intent".
Of what use is an artist sortname?

It would be my intent to use such a sortname as a way to lexicographically
sort artists in a single ordered presentation.
The actual presentation of an artist is the regular artist name but the sort
order is based soley the artist's single sortname.

What should such an order be?
It used to be that a well understood model for such sorting was a brick and
mortar music store or the ordering of one's physical media on the shelf.
Digital media opens up the possibility of indexing/sorting in many ways but
I think the intent of sortname should be to order in just one way -
lexicographically.
Is that the intent?
If so, we should say so.

I think it is good to use examples when discussing this.
Consider the following related artist names, all having "Les Claypool" as a primary member.

Sausage
Primus
Les Claypool and The Holy Mackerel
Oysterhead
The Les Claypool Frog Brigade
Colonel Claypool's Bucket of Bernie Brains
Les Claypool

If we were to have to choose a single order using a lexicographic sort of a single artist sort name, what would we like the result to be?
A record shop or my shelf might have all *Claypool* artists share the same sort name, say, "Claypool, Les" - effectively putting them all in the "Les Claypool" section.
In fact, my shelf has media sorted this way - effectively using "Claypool, Les" as the primary sort key and the release date as the secondary key.
Should this be our intent?
If so, we should say so.

Or should all *Claypool* artists have sortnames that all share a common root (say, "Claypool, Les") - effectively the primary key in a lexicgraphic sort - and have the rest of the artist name used as a secondary key.
Say:

Claypool, Les
Claypool, Les and Holy Mackerel, The
Claypool, Les Frog Brigade, The
Claypool, Les's Bucket of Bernie Brains

Chances are the latter ordering would present these artists contiguously (is this our intent?) but it wouldn't order them as they are on my shelf as the former ordering would.
Again, what is the intent?

If either is our intent, then there will, unfortunately, be some subjectivity involved.
This is unfortunate because it is much easier to describe, follow and enforce mechanical objective rules.

Note that I would (subjectively) choose artists Sausage, Primus and Oysterhead to have the sortname the same as the artist name.
Depending on one's intent, however, I can understand an argument for making the sortname of "Sausage" be "Primus".
See http://en.wikipedia.org/wiki/Primus_%28band%29#Band_history.

With regard to objective rules, I suggest a recursive, ordered, divide and conquer approach.
Divide the artist name into components, reorder the components and sort the components, recursively.

Consider "The Les Claypool Frog Brigade".
With regard to ordering, I would say the first division into components
would result in "The" and "Les Claypool Frog Brigade".
For sorting, the last component is more important than the first and since
we are doing a lexicographic sort it must appear first.
To suggest that a reordering has been done, a comma is inserted - resulting
in "Les Claypool Frog Brigade, The".
Then the components are addressed, recursively.
"The" is terminal so we are done but there is more work to do on "Les
Claypool Frog Brigade".
"Les Claypool" should be separated, ordered first and the other part(s)
ordered next.
Since there isn't a reordering operation here, there is no comma inserted.
We now have to sort "Les Claypool" and "Frog Brigade".
The latter is terminal but the first is not.
"Les Claypool" is divided into "Les" and "Claypool" and reordered - yielding
"Claypool, Les".
All components cannot be divided further so we end up with "Claypool, Les
Frog Brigade, The"

I have applied these rules on the other examples as well.
"Colonel Claypool's Bucket of Bernie Brains" is an interesting example
because "Claypool" was artificially expanded to "Les Claypool" to meet our
intent.
"Les Claypool's" is divided into "Les Claypool" and "'s", no reordering is
done but "Les Claypool" is divided an conquered - resulting in "Claypool,
Les's".

When ordering a name, I would think family name should have most precedence
(be ordered first), followed by the given name, the suffix (if any) and
prefix (if any).
I would consider Jr./Sr. as suffixes.

--
View this message in context: http://musicbrainz-mailing-lists.2986109.n2.nabble.com/RFC-320-Revised-Sortname-Style-tp6209460p6229502.html
Sent from the Style discussions mailing list archive at Nabble.com.

Message: 3
Date: Thu, 31 Mar 2011 21:08:09 -0700 (PDT)
From: rossetyler <rossetyler@gmail.com>
Subject: Re: [mb-style] RFC 320: Revised Sortname Style
To: musicbrainz-style@lists.musicbrainz.org
Message-ID: <1301630889974-6229657.post@n2.nabble.com>
Content-Type: text/plain; charset=us-ascii

More on intent...

Just as I think it is important to say what the intent is, I think it is
important to say what it is not.
I think that it is _not_ the intent of a sortname to be presentable.
The artist name is used for presentation, the sortname is used to order it.
Further, depending on what the intent is, the sortname need not be textually
related, in whole or part, to the artist name (take the Sausage and Colonel
Claypool's Bucket of Bernie Brains examples above).

--
View this message in context: http://musicbrainz-mailing-lists.2986109.n2.nabble.com/RFC-320-Revised-Sortname-Style-tp6209460p6229657.html
Sent from the Style discussions mailing list archive at Nabble.com.

_____
MusicBrainz-style mailing list
MusicBrainz-style@lists.musicbrainz.org
http://lists.musicbrainz.org/mailman/listinfo/musicbrainz-style

End of MusicBrainz-style Digest, Vol 72, Issue 1
****

08:16:01 <ijabz_> ijabz_ has joined #musicbrainz
08:20:11 <Dr-Hfuhruhurr> Dr-Hfuhruhurr has joined #musicbrainz
08:20:11 <Dr-Hfuhruhurr> Dr-Hfuhruhurr has joined #musicbrainz
08:30:44 <ocharles> ocharles has joined #musicbrainz
08:30:44 <nikki> nikki has joined #musicbrainz
08:36:37 <juhae> juhae has joined #musicbrainz
08:36:37 <Muz> Muz has joined #musicbrainz
08:36:37 <LotR> LotR has joined #musicbrainz
08:36:37 <enjayhch> enjayhch has joined #musicbrainz
08:36:37 <Sergey_Ivanov> Sergey_Ivanov has joined #musicbrainz
08:41:29 <CatCat> [01:28]   nikki   I wondered what would you do with
http://musicbrainz.org/release/d0f2e308-2bd6-4c2a-88d1-d78cc3049eaf.html
and works
08:41:29 <CatCat> [01:29]   nikki   the first two songs have the same
music, but the lyrics are a bit different
08:41:51 <CatCat> nikki, I guessed it before i saw it, I remember dat
version and bath version!
08:42:40 <CatCat> but in bath version there are more bubble sounds and in
date version there is sorta carnival/fairground backdrop
08:43:02 <CatCat> the karoke track I think was the bath version
spesifically.
08:43:12 <CatCat> * CatCat is really good with notecing things like this
08:47:02 <CatCat> but re cleanup, sorry about that, I was supposed to do a
trying, but apparently my wintervacation really meant "work on a bunch of
stuff that you've been putting of andomg dentists and omg stuf stuff stuff"
08:47:57 <CatCat> so I ask the scrollbakc readers and chatlogreaders this:
is Bruce Springsteen ok?
08:48:11 <CatCat> for cleanup o' the thing?
08:51:46 <CatCat> if noone objects strongly I'll make him the CotM and
start making the things tomorrow (today I have a dentis appointment, and
I'm still exhausted from yesterday :()
08:52:08 <reosarevok> Well, I know almost nothing about him, but I have
nothing against it either, specially if it is in a crappy state
08:52:50 <CatCat> I fear the list of nominations are a bit out of date, and
I can't very well just come up wioth soemthing personal heh
08:53:32 <CatCat> there are a few things missing
08:54:23 <ijabz_> ijabz_ has joined #musicbrainz
09:55:56 <ocharles> Why has http://musicbrainz.org/artist/a40423b7-
cea7-4e3d-85e8-d646601fbdf4.html not been deleted? It looks like an empty
artist to me
09:56:44 <ocharles> ah, track relationship
10:21:56 <JoeLlama> JoeLlama has joined #musicbrainz
10:22:58 <_bibi> _bibi has joined #musicbrainz
10:40:28 <Krystof> Krystof has joined #musicbrainz
11:56:42 <Tecfan> he misc'ed the shit out of Luka bloom
12:07:09 <Tecfan> http://andrewvos.com/2011/02/21/amount-of-profanity-in-

git-commit-messages-per-programming-language/
12:08:08 <pbryan> pbryan has joined #musicbrainz
12:08:10 <nikki> needs more perl
12:09:01 <_bibi> _bibi has left #musicbrainz
12:10:42 <ocharles> apparently we have 0 comments with "fuck" in
12:10:53 <ocharles> 5 with 'stupid' though :P
12:20:56 <Tecfan> ^^
12:36:45 <STalKer-X> STalKer-X has joined #musicbrainz
12:44:05 <Tecfan> if a label produces 500 cd's in a batch, will all these get the same disc id?
12:44:57 <nikki> I would imagine so, why?
12:45:06 <Tecfan> well I'm gonna make my own CD ;>
12:45:24 <Tecfan> with my own music
12:45:32 <nikki> (although there are still cases where windows generates one that's two seconds longer than the ones you'd get in osx/linux)
12:45:32 <Tecfan> and i would like them all to be identical
12:46:20 <Tecfan> so basically all i need to do is to make sure that every cd i burn has the exact same files burned to them?
12:46:31 <Tecfan> (in the same order and same offsets)
12:49:14 <Tecfan> all my material is originally in mp3.. but i am taking my time to find all the correct settings again and saving them to flac
12:49:22 <Tecfan> so i can get a bandcamp download ;>
12:49:29 <Tecfan> (with mbid's)
12:51:51 <nikki> * nikki grumbles at that redirecting on exact match thing
13:20:33 <pronik``> pronik`` has joined #musicbrainz
14:05:47 <DarkerAudit> DarkerAudit has joined #musicbrainz
14:24:50 <Vorpal> Vorpal has joined #musicbrainz
14:30:26 <lugo> lugo has joined #musicbrainz
14:50:07 <kurtjx> kurtjx has joined #musicbrainz
14:50:19 <jesus2099> jesus2099 has joined #musicbrainz
14:51:22 <jesus2099> nikki: I don't see the problem with your karaoke that is a karaoke for two songs, just don't put the lyrics AR on the karaoke track if they are different : http://musicbrainz.org/mod/search/results.html?minid=14097224&maxid=14097225
14:57:05 <jesus2099> catcat: oh so there is a different music and the karaoke one applies to one of them.. I could cancel that "date" edit then ...
15:01:44 <caller_6> caller_6 has joined #musicbrainz
15:17:12 <_bibi> _bibi has joined #musicbrainz
15:17:23 <_bibi> _bibi has left #musicbrainz
15:27:22 <pbryan> pbryan has left #musicbrainz
15:30:24 <zazi> zazi has joined #musicbrainz
15:51:05 <__lucas> __lucas has joined #musicbrainz
15:53:33 <ocharles> ocharles has joined #musicbrainz
16:00:03 <ocharles> ocharles has joined #musicbrainz
16:09:28 <hawke_> hawke_ has joined #musicbrainz

16:42:30 <caller_6> anybody up for a legalistic (c) question?
17:16:50 <Tecfan> just ask
17:18:48 <caller_6> I suppose you're right...
17:22:31 <caller_6> I found an instance where somebody had quoted a
CreativeCommons source in an annotation. The editor /did/ give attribution,
so normally that'd be fine, but is that true when the annotation itself is
in the public domain? (at least I /think/ annotations are public domain)
17:59:20 <ruaok> ruaok has joined #musicbrainz
18:06:33 <caller_6> ruaok:ping
18:08:36 <caller_6> Are annotations in the public domain?
18:11:31 <nikki> no
18:11:50 <ruaok> pong
18:13:27 <CatCat> ruaok: ping
18:13:46 <CatCat> since no one has argued, Bruce Springsteen is the new
cotm
18:14:31 <ruaok> pong
18:14:32 <caller_6> nikki: annotation are CC then?
18:14:33 <ruaok> ok
18:14:40 <ruaok> caller_6: yes
18:15:02 <CatCat> I am exaustd today so I'll fixor everything tomorrow
mornign :)
18:15:04 <CatCat> arg
18:15:15 <caller_6> * caller_6 got confused in the crossfire
18:15:18 <CatCat> dentists and godman busses that wont even let me in wtf
18:19:33 <CatCat> I was os mad, I wanted to ragequit. i was so mad I left
the preorder tikkit for PkMn B&W at the dumb library. wtf arg
18:20:18 <Tecfan> * Tecfan fell off at tikkit pkmn bw
18:20:52 <CatCat> hva mener du?
18:23:37 <Tecfan> hva i all verden er PkMn?
18:23:39 <Tecfan> pokemon?
18:23:50 <CatCat> ja hva ellers?
18:24:07 <Tecfan> trodde tikkit var en matrett :P
18:24:10 <CatCat> den forkortelsen er jp brukt i spillene ogs√●
18:24:37 <CatCat> det er en herpa m√●te √● skrive "ticket"
18:24:38 <CatCat> p√●
18:28:03 <Tecfan> hm
18:28:15 <Tecfan> spiller du pÔøΩ color eller? ;p
18:30:40 <CatCat> om jeg gj√∏r hva da?
18:31:32 <Tecfan> nvm
18:32:06 <CatCat> n√● gj√∏r du jp det selv, bruker forkortelser. :P
18:34:18 <Tecfan> pssh, :D
18:34:25 <CatCat> re wikis: can any wikizenguru guys or gals make a
"category: CotM" template?
18:35:08 <nikki> what do you mean?
18:35:15 <nikki> you don't need a template just to add a category..?
18:38:52 <CatCat> i mean a short page that say The CleanupOfTheMonth for

blah is fobar
18:39:06 <CatCat> like http://wiki.musicbrainz.org/Cleanup_Of_The_Month/Benny_Goodman
18:39:18 <CatCat> so I wont forget to use categories and stuff
18:40:25 <PasNox> PasNox has joined #musicbrainz
18:50:00 <CatCat> wikigurus: is there a way to MAKE http://wiki.musicbrainz.org/Community_Project the categorypage for http://wiki.musicbrainz.org/?title=Category:Community_Project complete with info under each thing?
18:50:05 <PasNox> PasNox has joined #musicbrainz
19:03:22 <ijabz> ijabz has joined #musicbrainz
19:29:58 <v6lur> v6lur has joined #musicbrainz
19:32:43 <ruaok> ruaok has joined #musicbrainz
19:36:25 <Tecfan> grats)
19:38:58 <ruaok> thx
20:00:05 <PasNox> PasNox has joined #musicbrainz
20:05:17 <__lucas> __lucas has joined #musicbrainz
20:07:45 <ocharles_> ocharles_ has joined #musicbrainz
20:18:18 <Krystof> Krystof has joined #musicbrainz
20:19:55 <Dremora> Dremora has joined #musicbrainz
20:23:59 <CatCat> disregard what I said re: tomorrow, the pages for Bruce Springsteen is up, anyone mind making a forum topic?
20:32:39 <caller_6> CatCat: forum topic to announce the CotM?
20:33:30 <caller_6> yeah, I'll post it
20:54:16 <reosarevok> reosarevok has joined #musicbrainz
21:03:47 <PasNox> PasNox has joined #musicbrainz
21:06:12 <hawke_> hawke_ has joined #musicbrainz
21:10:31 <ijabz> ijabz has joined #musicbrainz
21:19:03 <DarkAudit> DarkAudit has joined #musicbrainz
21:23:11 <Tykling> Tykling has joined #musicbrainz
21:44:29 <frewsxcv_> frewsxcv_ has joined #musicbrainz
21:52:46 <MBChatLogger> MBChatLogger has joined #musicbrainz
21:52:46 <MBChatLogger> is logging
21:52:46 <niven.freenode.net> Users on #musicbrainz: MBChatLogger Dremora reosarevok kuenx
22:15:46 <DarkAudit> DarkAudit has joined #musicbrainz
22:37:43 <MBChatLogger> MBChatLogger has joined #musicbrainz
22:37:43 <MBChatLogger> is logging
22:37:43 <brown.freenode.net> Users on #musicbrainz: MBChatLogger KingJ yvesr tenthirteen czaja MClemo creature Mineo CatCat ianmcorvidae Clint sezuan flamingspinach gnu_andrew luks srotta evilmoo praest76 navap Knio FauxFaux| murdos 5EXAB9C3F FauxFaux Tecfan chrisb Jormangeud ddaydj hangy nikki juhae Muz LotR enjayhch Sergey_Ivanov STalKer-X lugo kurtjx caller_6 ruaok __lucas ocharles_ Krystof PasNox hawke_ ijabz DarkAudit Tykling frewsxcv_ v6lur Vorpal kepstin Rondom warp VxJasonxV stuffcorpse
22:37:43 <brown.freenode.net> Users on #musicbrainz: Tecfan_work