MusicBrainz Non-Profit White Paper

by Robert Kaye
rob@musicbrainz.org
February 2003

Table of Contents

1. Introduction

MusicBrainz aims to create a music information commons where the community creates and maintains a public database of information about music. This music metadata will enable non-ambiguous communication about music, and will allow the Internet community to discover new music without any of the bias introduced by marketing departments of the recording industry.

The MusicBrainz project has been around since the fall of 1998 (previously the CD Index), and is now gathering more support from the community and partnering companies. In order to give MusicBrainz some legal muscle and to ensure the future availability of the dataset, it is proposed that MB be incorporated in California as a non-profit corporation.

However, creating and running a non-profit corporation costs money, and with limited resources, MusicBrainz will depend on donations from the community and industry sponsors to elevate it to the next step.

2. MusicBrainz Today

The first version of MusicBrainz, which nears completion during the first quarter of 2003, focuses on creating an open database of basic music metadata which can be used for identifying audio CDs and digital audio tracks (MP3, Ogg/Vorbis, WAV, etc.). MusicBrainz is comprised of three separate components which all work together to enable users to semi-automatically identify music and apply clean metadata tags to their music collection:

1. *MB Web site:* The MusicBrainz web site allows anyone on the net to search, browse, and maintain the community metadatabase. The web site users (moderators) can add new metadata to the site, edit or correct existing metadata, and delete incorrect metadata via the web-based moderation system.

2. *MB web service/client library*: All of the MusicBrainz data is available to the public via the RDF-based web service. A web service client can search for and request information about any artist, album or track in the database. A client library released under the LGPL is available for developers who would like to support MusicBrainz in their application. This client library abstracts out the details of interacting with the MusicBrainz web service, and allows a client developer to add metadata lookup to their applications in a short period of time.

3. *MB Tagger*: This 32 bit Windows application (similar applications with support for other platforms are also in development) takes an end-user's collection of MP3, WAV and Ogg/Vorbis files, generates an acoustic fingerprint (TRM Id) for each track and, using the fingerprint, looks up the track metadata. If the main server does not have the metadata available, the application guides the user through the process of entering the missing information into MusicBrainz so that future users may benefit from the new metadata. After the proper metadata has been downloaded/entered, new metadata tags are written to the user's audio files.

The basic metadata includes a list of artists and artist aliases (e.g. alter-ego names, alternate band names and common abbreviations) and for each artist a list of albums and the tracks for each album. MusicBrainz assigns each artist, album and track a unique identifier, which can be used to refer to a particular artist/album/track without having to deal with the semantics of correct spelling and conflicting names in the database.

These identifiers provide the Internet community with a means to establish a meaningful computer-based dialog about music. This unambiguous dialog is enabled by an RDF based web service interface and presents the first baby steps towards the "Semantic Web", where computers can carry on a meaningful discussion without involving human beings. The RDF used in the web service uses portions of the Dublin Core and is documented on the MusicBrainz site. MusicBrainz encourages others to use the RDF in other future music applications to enable a host of new applications and features that are not possible today.

For instance, it is not possible today to exchange a playlist with a friend, since your friend may not have the same files that you do; even if your friend does, the files may be located in a different location on the hard drive. Using MusicBrainz, a user can create a playlist that consists solely of MusicBrainz track identifiers, and then send that playlist to their friend. Their friend will be able to feed the playlist to their MusicBrainz-enabled audio player and then have the player match up the available tracks. If some of the tracks are not available in the collection, the player could go out to music sites such as EMusic.com, MusicNet or Pressplay to download the missing tracks. The MusicBrainz identifiers allow future audio applications to carry on unambiguous conversations about music and to enable a whole new set of features for music enjoyment and music discovery.

The MusicBrainz dataset has been created and maintained by its user base of over 2000 volunteers. Since its inception as the CD Index in the fall of 1998, and the consequent renaming to MusicBrainz in the fall of 2000, the database has seen more than 160,000 additions and changes (moderations) to the database. Even without any promotion of the site, and all of the software just now emerging from a beta state, the dataset is growing and improving in quality. To see the latest statistics on MusicBrainz, please visit: http://musicbrainz.org/stats.html.

MusicBrainz's human moderation approach encourages participation in the data maintenance process and thus yields higher quality data, since many eyes will spot even the smallest mistakes. Active moderation, concise technology for identifying music and a carefully designed database allows MusicBrainz to collect data with greater accuracy than services like GraceNote. The GraceNote service suffers from an overwhelming number of errors and duplicate entries in their database, and without a focus to reduce duplicates and to correct errors in the database, they cannot compete with MusicBrainz in the long run.

Furthermore, GraceNote charges serious amounts of money for severely restricted access to its data. FreeDB, the free alternative to GraceNote, has not created any new technology to advance the state of the project. FreeDB's goal is to provide a service that is free and backward compatible to the old GraceNote/CDDB service. This gives MusicBrainz the advantage to create the first well-edited, highly structured and comprehensive music encyclopedia on the net.

Once the TRM (acoustic fingerprint) and audio CD based music identification portions of MusicBrainz have been completed, the service is poised for a significant increase in the number of users contributing to and using MusicBrainz. This will provide a powerful alternative resource for non-commercial music developers and a very low cost alternative for commercial music services and channels.

3. MusicBrainz Tomorrow

The basic metadata framework that the first generation of MusicBrainz puts into place will enable more comprehensive and subjective metadata to be added to the community metadatabase. A few possible additions include:

1. *Reviews/biographies/ratings*: Unlike the rest of the existing MusicBrainz dataset, artist/album reviews/ratings and artist biographies and are not factual metadata, and thus they will require a different approach in collecting and maintaining. However, this subjective metadata may present the most significant revenue source for MusicBrainz. (see below for details)

2. *Music Discovery*: The advanced music classification from above will allow MusicBrainz users to browse the available genres and discover new music as they find genres that describe their own musical tastes. Combining the music classification with user-contributed information about their own musical collections will enable MusicBrainz to offer collaborative filtering services to its users.

3. *Advanced Music Classification*: Today's music classification systems leave a lot to be desired, since music classification is a highly subjective task, and few subjective systems have been developed to date. However, MusicBrainz can harness the power of many users to create a representative classification system that will evolve over time as musical genres evolve. Using data collected from thousands of users will enable MusicBrainz to statistically infer Genre Curves for artists and albums.

4. *Detailed Music Information*: MusicBrainz will expand to cover more information about music such as artist web pages, official fan web pages, detailed support for classical music (e.g. composer, opus number, orchestra, conductor, etc.), and any other relevant pieces of information that will make MusicBrainz into a comprehensive music encyclopedia.

5. *Music Genealogy*: MusicBrainz may keep track of which artists/performers/engineers contributed to a piece of music, and when these contributions took place. Combining this contribution data with data on how artists influenced each other will create a genealogy of modern music. Imagine being able to track Britney Spears back to Beethoven!

These are just a few of the possible future directions of MusicBrainz. The actual directions will be heavily influenced by the MusicBrainz partners/sponsors to create a mutually beneficial relationship between MusicBrainz and its partners and sponsors.

4. MusicBrainz Licenses

MusicBrainz is devoted to using the right licenses for the right job and thus the GPL (GNU's General Public License) is used for the server software and the LGPL (GNU's Lesser General Public License) for the client library. The use of the LGPL allows even closed source applications to use the client library to access the MusicBrainz server.

The overall goal is to remove as many obstacles to accessing the MusicBrainz dataset as possible and to foster the inclusion of MusicBrainz technology in third party applications. To support this goal, MusicBrainz makes the dataset available to the public by placing portions of the dataset into the Public Domain and releasing other portions under Creative Commons' Attribution-NonCommercial-ShareAlike License 1.0:

1. *Core data*: The core data is comprised of the artist, artist alias, album, and track information, as well as the CD Index identifiers, and TRM identifiers. All of this data is released into the Public Domain.

2. *Derived data*: The derived data consists of artist, album and track text indexes, as well as moderation and voting information, which is released under the Attrbution-NonCommercial-ShareAlike License from the Creative Commons.

3. *Subjective data*: In the future MusicBrainz will collect artist biographies, album reviews, music ratings, and other non-factual data and also release them under the Attrbution-NonCommercial-ShareAlike License.

To some people the use of the Public Domain for the core data may come as a surprise. However, the United States Supreme Court decided that facts are not copyrightable and all of our core data is essentially comprised of facts. This limitation, combined with the desire to have commercial enterprises use the MusicBrainz core data to extend the reach of this data, makes the Public Domain a perfect choice.

5. MusicBrainz and Commercial Enterprises

Even though MusicBrainz is an open source and open data project, MusicBrainz actively encourages companies to participate in the MusicBrainz community. The availability of the core dataset in the Public Domain encourages companies to work with and link to the MusicBrainz dataset without having to navigate a complex maze of license requirements.

MusicBrainz is not hostile towards commercial (for-profit) corporations! On the contrary -- MusicBrainz will only reach its full potential if commercial corporations use the dataset and encourage their customers to participate in the MusicBrainz community. Any and all corporations around the globe are encouraged to use the MusicBrainz core dataset to establish meaningful and non-ambiguous conversations about music.

The derived and subjective data components in MusicBrainz are licensed under the Creative Commons Attribution-NonCommercial-ShareAlike License, which prohibits the use of the data in a commercial setting. However, MusicBrainz will make commercial licenses to the data available to companies that wish to use the data in a commercial setting. The income from these license agreements will provide MusicBrainz with the needed revenue to ensure that the dataset continues to evolve and remains available to the public.

However, many companies are skeptical about using open source software because there is no one to call (or hold responsible) should the software fail. Open data projects like MusicBrainz are in a similar position -- what if the data is wrong? Or not in the database at all? The answer to this lies in the MusicBrainz community -- the community is comprised of individual contributors who work hard to enter and correct the data in the system. The MusicBrainz server software also enforces a peer review system, under which users must review and approve changes made by other users. The peer review system combined with the motivation, expertise and pride of its contributors will ensure that the data in MusicBrainz will be comprehensive and reasonably correct.

Only reasonably correct? No one can guarantee that all the data in a database is correct. Not even the commercial companies that provide metadata services can give this assurance. The MusicBrainz community will respond to problems found in the database and fix mistakes faster than any commercial company with paid contributors can, since the MusicBrainz community is global and is never closed for business. Furthermore, the community is more supportive of MusicBrainz than of other commercial services due to its open nature.

Another area corporations are skeptical about is the issue of service reliability. The MusicBrainz servers have always lived in professional colocation facilities with excellent connections to the Internet, and even though there has not been a legal corporation watching over the servers for the first four years of its life, MusicBrainz has had only a handful of minor service interruptions.

In the future, MusicBrainz plans to create a network of mirror servers that will mirror the dataset across the globe. Any corporations that would like to work with MusicBrainz, but would prefer to handle their own servers for reliability and added load balancing, will be welcome to operate their own MusicBrainz mirror server. This option leaves all the service reliability concerns in the hands of the corporation.

6. MusicBrainz Non-Profit Corporation

In order to ensure that the MusicBrainz dataset will continue to exist and continue to be available to the public, a tax-exempt non-profit corporation (503.c.3) should be created. This non-profit should adopt a set of bylaws which will state that MusicBrainz will make all metadata created by the MusicBrainz community available to anyone who wishes to download the data. The MusicBrainz corporation should consider itself the guardian of the MusicBrainz dataset and its community, and should take the necessary actions to ensure that MusicBrainz can continue its mission.

The MusicBrainz non-profit should strive to become self sufficient over the course of 2-3 years. To achieve this independence, it should pursue the following possible revenue streams:

1. *Contributions from the community*: Users of the MusicBrainz Tagger will greatly benefit from the project by having the tagger automatically clean up the metadata present in a user's collection. For-profit companies charge for this service, and MusicBrainz should ask users for a $10 contribution for the service of cleaning up the metadata.

2. Google style ad-words program: As MusicBrainz gains more users, it will be possible to offer an ad-words program similar to the one pioneered by Google. Third parties will be able to purchase small and unobtrusive advertisements that will be shown on artist/album pages.

3. *License artist/album reviews and biographies*: When MusicBrainz provides the infrastructure to collect and manage album/artist reviews and biographies, it will ask the authors of these works to assign the copyright to MusicBrainz. These reviews and biographies will then be made available to the public under the Creative Commons Attribution-NonCommercial-ShareAlike License. Furthermore, as this collection of reviews and biographies becomes comprehensive, MusicBrainz will offer a commercial license to this content for use in commercial applications and web sites.

4. *Provide MusicBrainz dataset services*: As corporations switch away from proprietary music metadata services, MusicBrainz will gain a larger user base. However, since MusicBrainz is community funded it will be unable to provide the bandwidth for millions of users to access the dataset. Large commercial customers will be encouraged to setup their own MusicBrainz mirror servers to handle the load of their own customers. However, some commercial customers will not want to deal with this in-house and would rather contract out these hosting and integration services. MusicBrainz will be available for hire to carry out the hosting and integration of the dataset on behalf of corporations. In the same spirit, if commercial customers would like to have a dedicated support staff for addressing problems with the service or data, MusicBrainz will also be able to provide these services.

The above revenue streams will take some time to develop, but over time MusicBrainz will strive to grow its revenue and become self sufficient. Should MusicBrainz find itself in a position of having excess revenue (where a for-profit company would pay a dividend), it will offer grants or awards to open source/open data/music projects and their developers.

MusicBrainz has never had anything to hide, and all of its business has been visible to the public. The finances are transparent and all discussions are carried out in a public forum. With this approach, MusicBrainz will attempt to create a new kind of non-profit corporation that can continue to hold the trust of its community.

7. The Road Towards the MusicBrainz Non-Profit Corporation

Unless a generous donor donates a large sum money, MusicBrainz will need to take several steps to become a fully fledged non-profit corporation:

Step 1: Form the non-profit corporation

The first step incorporates the non-profit corporation, sets up all the needed legal paperwork and prepares the non-profit to exist for at least one year. The estimated cost of incorporation by a lawyer is about $1,500, and the cost of paying the rent and DSL for the office that has been used as the MusicBrainz headquarters in San Luis Obispo is about $6,000/year. Total: $7,500

Step 2: Sponsor the first full time employee

The second step will sponsor the employment of Robert Kaye for one year at a cost of $60,000 - $75,000/year. Alternatively, any corporation interested in sponsoring MusicBrainz could hire Robert Kaye to continue working on MusicBrainz.

Step 3: Apply for grants

The goal of the third step is to apply for a number of grants that will provide funding for MusicBrainz for the next 2-3 years while it becomes self-sufficient. During this phase, MusicBrainz will start to develop the above mentioned revenue streams.

Step 4: Become self-sufficient

After having been awarded a grant to provide for the operating expenses for 2-3 people for a duration of 2-3 years, MusicBrainz will aim to further refine its revenue streams while it continues to develop its core technologies.

8. Conclusion

Community feedback about the MusicBrainz project has been overwhelmingly positive; now is the right time to take MusicBrainz to the next level and create a non-profit corporation. If you believe that MusicBrainz has the power to make a difference, please consider contributing money to MusicBrainz. While we are looking for sponsors to contribute larger donations, we welcome any donations. Anything helps to move the project forward and keep it alive.

If you are interested in sponsoring or supporting MusicBrainz, please contact rob@musicbrainz.org. If you feel like making a smaller donation to MusicBrainz, please use PayPal to donate to donations@musicbrainz.org. If you think that MusicBrainz is a good project but cannot support it yourself, please pass this document on to someone who may be able to help.