MusicBrainz and Its Peers: Comparing Cultural Commons

Brian W. Carver[†]

09/12/11 Draft

Abstract

In 2011, Jess Hemerly completed a thorough empirical study of MusicBrainz, a metadatabase of structured information about musical releases maintained by a volunteer online community. This paper builds on her research by comparing and contrasting her findings regarding MusicBrainz with the findings of scholars that have studied open source and free software projects and other instances of online peer-production such as Wikipedia. I begin the process of identifying what makes some of these efforts succeed and what may be missing when other similar efforts fail, working towards a theory of successful online peer-produced information commons.

Introduction

Why do people contribute their time, skills, and knowledge to an information commons when often the direct benefits to them personally are typically abstract and intangible? When researchers observed this phenomenon in free software projects it was possible to attribute some of the motivation to reputation and skill-enhancing activity that often might lead to the more tangible monetary rewards of a better job. When Wikipedia became large enough and endured long enough that it could no longer be ignored, researchers had to acknowledge that most people do not get better jobs because of their excellent Wikipedia-editing skills and thus other motivations had to be taken seriously and explored. Many other attempts at internet-based collaboration on the production of an information commons have had setbacks or failed. Other more productive collaborations, though obtaining some measure of success, remained small niche efforts that made it possible to argue that these should be discounted as uninteresting or unimportant. Thus some have been tempted to conclude that there is simply something unique about the development of software and encyclopedias that makes these online collaborative efforts succeed and others fail. To counter such a hypothesis, additional examples of successful internet-based collaboratively-built information commons are essential. MusicBrainz is such an example.

---

† Assistant Professor, University of California Berkeley School of Information. Special thanks to Jessica Hemerly, whose ground-breaking research on the MusicBrainz community made this follow-on paper possible.

As different examples of commons-based peer production are compared and contrasted, the hope is that theories can be developed to explain the factors that make some successful and others fail. With such theories in hand, the further hope is that online commons-based peer production efforts could be designed from the outset in ways likely to increase the probability of their success.

What is MusicBrainz?

MusicBrainz is a metadatabase of structured information about musical releases maintained by a volunteer online community. Users of the MusicBrainz website can contribute information about musical artists, releases, tracks, and other information about musicians and their works. When one places an audio CD in a computer or uses a software-based audio player to play a digital audio recording, one typically makes use of a music metadatabase to provide the software with information about the recording's title, artist, length, track number, etc. The MusicBrainz database contains such music metadata..

Researching MusicBrainz

In May 2011, Jess Hemerly completed a nearly 18-month study of MusicBrainz. (Hemerly, 2011). She engaged in participant observation, administered a survey to obtain quantitative results, conducted numerous interviews, and used additional means of direct investigation in completing her research. Her results will be referenced repeatedly in what follows.

Is MusicBrainz Successful?

Before comparing MusicBrainz with other cultural commons in the hopes of eliciting commonalities that might guide us towards the factors likely to yield a successful cultural commons, we must ask whether MusicBrainz is itself a success. Both free software projects and Wikipedia faced early doubts about their reliability and quality, but as these efforts have grown, persisted, and improved, most such doubts have been answered by the continued success of such projects. In what does this “success” consist?

[Note: This section will likely be revised to adapt to MusicBrainz some of the potential indicators of success from Kevin Crowston, James Howison, and Hala Annabi , *Information Systems Success in Free and Open Source Software Development: Theory and Measures*, Softw. Process Improve. Pract. 2006; 11: 123–148 .]

Is MusicBrainz meeting its own stated aims?

We might define “success” as meeting one's stated goals. The MusicBrainz

General FAQ says many things about its aims. It "is intended to be a free, online encyclopedia of music information...containing all the information you would ever want to know about songs, releases, and artists." MusicBrainz also contrasts its aim with that of commercial efforts, aiming to provide no-cost access to use the database in one's music player and the ability to freely download large parts of the database. MusicBrainz also contrasts its database structure with that of a no-cost competitor, FreeDB, by expressing a preference for a database grouped by artists, that avoids duplicate entries for the same releases, and that permits a single track to appear on multiple releases. That is, the maintainers of MusicBrainz have a notion of database quality that they seek to achieve and that they believe competing efforts lack.

As to each of these aims, MusicBrainz does quite well. As of September 2011 it contains information on over 619,000 artists, 965,000 releases, and 10 million recordings. The commercial entity Gracenote reports coverage of "more than 400,000 artists" though it claims significantly more releases and recordings. (Gracenote, 2010). MusicBrainz' coverage of an artist typically includes all albums, singles, compilations, soundtracks, and live recordings. It also includes the dates of these releases, the format, label, number of tracks, country, catalog number, and even barcode. The length of each track is provided, as are the names of the lyricist, writer, and composer. When a work has been recorded by multiple artists, this information is also conveniently displayed. Through this information provided directly and by providing links to external sites such as Wikipedia, MusicBrainz is at least making significant progress toward providing "all the information you would ever want to know" about a given song, release, or artist.

MusicBrainz also is achieving its aim of making its database available for use at no cost, and provides frequent database dumps at no cost for those who might wish to download the database.

The above list of data elements that the MusicBrainz database contains also indicates that MusicBrainz is achieving its aim of providing a structured database that makes it possible, for example, for a track to appear on multiple releases.

Is MusicBrainz sustainable?

Another measure of success might be defined as having and maintaining the resources one needs to do the work one seeks to do. On that front, there are several indicators that MusicBrainz is sustainable across several types of needed resources: technical, financial, personnel, and legal.

First, the technical requirements to maintain the project are relatively low. MusicBrainz relies on the free software database PostgreSQL, which can run on

commodity hardware with a free software operating system. The entire set of databases, including edits, votes, annotations, editor data, and statistics amounts to less than 3 gigabytes total and so can be completely replicated on a wide variety of inexpensive storage mediums.

The financial resources of the project are provided by charging external music services, including Last.fm and BBC Music, for access to hourly updates of the live data feed, (Hemerly, 2011), by various affiliate programs, and by donations from individuals and corporations. A hardware fundraiser to raise $15,000 was successfully completed in March 2011. (MusicBrainz Blog, 2011). MusicBrainz is operated by the non-profit MetaBrainz Foundation, which publishes its finances on a monthly basis. The MetaBrainz Foundation's income has slightly exceeded its expenses every year since its inception in 2004. (MetaBrainz, n.d.). The MetaBrainz Foundation is able to provide modest salaries to its Executive Director and to a few MusicBrainz developers. The income from the live data feed has in recent years been a relatively large percentage of overall income and thus the loss of such income would require either a scaling back of expenditures or a significant increase in other revenues. However, many free software projects continue for years without any such income at all, so it would seem odd to argue that the MetaBrainz Foundation should have greater diversity in its income sources, when in fact it is uniquely privileged to have such an income stream.

MusicBrainz would stagnate without active "editors" or contributors to continually add new releases to the database and to further curate existing entries.

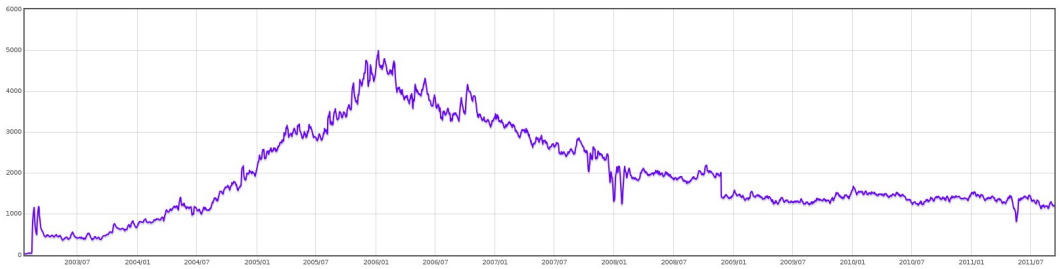

Figure [x]: "Active Editors" of MusicBrainz (approx. July 2003 – July 2011)

While the number of active editors has fallen off of the peak in 2006, it appears to have stabilized in late 2008 and remained fairly consistent since that time, moving between approximately 1200 to 1500 active editors at any given time. It is difficult to say whether this number of active editors is sufficient to the task.

Wikipedia also experienced exponential growth in the number of its editors until 2007 and has since reached and maintained a fairly constant number. (Suh, 2009). It remains to be seen whether these plateaus represent a certain maturity level

within a project and can be maintained indefinitely or whether they are part of a larger pattern that will see subsequent increases or decreases in participation.

MusicBrainz also exhibits what I will call "legal" sustainability. This is particularly important to MusicBrainz because of its history. MusicBrainz was launched in response to Gracenote taking CDDB private. CDDB was an earlier music metadatabase that had been made by volunteers and many of these volunteers were angry when their contributions were appropriated for commercial use. MusicBrainz has taken several steps to ensure that this could never happen to MusicBrainz and to ensure that even if the MetaBrainz Foundation ceased functioning, others could continue the project or, if necessary, pursue a forked version of the project. This is enabled by the fact that the core database is dedicated to the public domain and regular snapshots are made available for download. Additional portions of the database, such as annotations and tags, are made available under a Creative Commons Attribution-NonCommercial-ShareAlike 2.0 license, which would allow anyone willing to offer the information on these same terms to do so non-commercially by attributing MusicBrainz. The MusicBrainz server software is also licensed under the GNU General Public License, and thus in the event it was necessary, there would be no legal barrier to someone setting up a complete mirror of both the MusicBrainz server and its database, providing another type of sustainability. Each of these deliberate choices is also memorialized in the MusicBrainz Social Contract.

Success as Usefulness or Unique Value

Finally, one might define success as being "useful" to someone or providing some "unique value" to someone. That is, if no one or very few people used the results of the project, and if no one found it to provide at least some value that differentiated it from projects that provided for the same uses, then it would be hard to characterize such a project as successful. Here again, MusicBrainz fares well. The MetaBrainz Foundation's 2009 annual report showed that hits to musicbrainz.org were at approximately 50 million hits annually. There continue to be active editors and various music-related software programs that seek to make use of the database daily.

In sum, MusicBrainz is a successful example of commons-based peer production. In the next section, its features will be compared with other such successes in an attempt to tease out lessons for future efforts.

Comparing MusicBrainz to Other Cultural Commons

Structural Features

Modularity

MusicBrainz is an information commons and the information in the MusicBrainz database can be entered at a very fine level of granularity. That is, if you find that an artist's MusicBrainz page does not include a link to their entry on Wikipedia, then you can add just that URL. Thus, the task of editing the MusicBrainz database is exceedingly modular. The modularization of tasks in many free software projects[1] is believed to lower barriers to contribution (von Krogh, 2003) and allow individuals to work on tasks independently. (Lerner, 2002). Wikipedia shares this modularity in that one can edit an entry to add or remove a single character. In one way, MusicBrainz is more modular, because its database schema almost completely determines the range of acceptable contributions. There may be several ways to fix a software bug or improve a Wikipedia article, but the MusicBrainz database has a predetermined set of fields. Since one of the benefits of modularity was supposed to be that participants could contribute with less hierarchical direction (Tapscott and Williams, 2006), limiting the range of permissible types of contributions may even further decrease the need for direction.[2]

Plausible Promise

This modularization may also contribute to what Eric Raymond called the "plausible promise" of a piece of software (Raymond, 1997) and what Ostrom has referred to as the resource attribute of "feasible improvement." (Ostrom, 1998). Users find the MusicBrainz database in an already useful state, but see some small, feasible contribution that they could make which would improve it. Therein lies the "plausible promise" of an even better database.

Condition Indicators, the Predictability of Availability, and User Autonomy

Much like free software projects that provide their source code in a live online version control system that is browsable and retrievable by the public at any time, the MusicBrainz database is available for immediate search and inspection by any visitor to the website. The proposed changes to the database, the open edits, are also publicly browsable. In terms of access to their underlying content, both MusicBrainz and Wikipedia provide full access to the public at all times. This

1 It should be noted that features of free software projects generally will occasionally be described, but this is not meant to imply that they are all uniform.

2 The idea that peer production is a leaderless anarchy has been over-stated. On the contrary, [cite all the counter-examples].

ease of autonomous access to the underlying resources can provide reliable and valid indicators of the condition of the resource system at a relatively low cost and contributes to the predictability of the availability of such resources, which are nearly always available. Such access and retrieval capabilities are available without seeking specific permission from project leaders. Ostrom has identified such factors as important resource and appropriator attributes in natural resource commons. (Ostrom, 1998).

Resource Salience

Users of free software programs can in some cases become dependent upon that software for a major portion of their business or livelihood. This is not to say that alternatives do not typically exist for such users, but in many cases switching costs would be high. It seems unlikely that many users of Wikipedia are dependent upon it for their livelihood. It is an immensely convenient source of information online, but Wikipedia does not seek to be a business-critical resource and it seems likely that acceptable substitutes could be found for most queries, even if more difficult to acquire. MusicBrainz appears to sit somewhere in between these two. For most users, the metadata in MusicBrainz assists them in organizing their personal music collections. This is something these users may take seriously, but it rarely affects their livelihood. However, at least BBC Music and Last.fm, those who are paying for access to the live database feed, have become dependent upon MusicBrainz for such information and would have to seek an alternative provider in its absence. In cases such as these the users of the resource have an interest in the resource's well-being and continued existence. Even where a user's livelihood is not dependent upon the resource, so long as they expect to be a repeat user of the resource, then their desire to accomplish the tasks that the resource assists them in doing can act as a motivator to see to the resource's well-being and continued existence.

Governance

MusicBrainz largely relies on consensus-based decision-making. (Hemerly, 2011). Through the MusicBrainz-style mailing list, decisions are made about proposals to modify the Style Guidelines, which detail things such as how to capitalize or abbreviate titles. Major changes to the Style Guidelines occur through a two-step process on the mailing list. First, a user proposes a change through a request for comments (RFC). If the comments are mostly positive, then the proposal moves to a Request for Veto (RFV), where any member of the mailing list can veto the change. Those not vetoed become new guidelines. (Hemerly, 2011).

Changes to the database itself are resolved by voting by registered members. Users that have established a history of accepted contributions can vote on whether to accept any open edit. There are a large number of edits and some are

not voted on at all. Edits that receive no votes are added to the database after 14 days. In that time, some users have typically reviewed the edit, and found nothing so clearly objectionable in it to cause them to vote no, but also they may lack sufficient confidence in an edit's accuracy to vote affirmatively. Wikipedia allows a user's edits to become live on the website immediately and votes are generally eschewed, a difference in approach from MusicBrainz. However, the wiki software makes tracking the history of a page and doing reversions part of its very nature. With such tools, a bolder approach to accepting new edits becomes more feasible.

Free software projects are not uniform in their governance structures, but making decisions through an organized process of reaching consensus is certainly among the strategies employed.

Conflict-Resolution Mechanisms

[add]

Contributor Demographics

Gender

When Ghosh reported that 98.8% of respondents to his FLOSS developer survey were male (Ghosh, 2005), this was not an unexpected result and he noted that it was in accord with the 98.6% reported in the larger WIDI survey (Robles-Martínez et al., 2001) and the 97.5% reported as male in the Boston Consulting Group survey (Lakhani and Wolf, 2005). Ghosh concludes that even accounting for various sources of bias it is unlikely that female participation in the FLOSS developer community is much higher than 5–7 percent. This imbalance has sometimes been attributed to the gender imbalance in computer science generally. In 2002, Margolis and Fisher reported that at top research universities about 15 to 20 percent of computer science majors are female, and in advanced-placement computer science exams at the high school level, it was only 15 percent women. (Margolis and Fisher, 2002).

Research on Wikipedia has found that only 16.1% of the editors who started editing Wikipedia in 2009 were female and that the gender gap is more pronounced when looking at high-activity editors. (Lam et al, forthcoming). This was roughly consistent with the earlier findings of a survey commissioned by the Wikimedia Foundation, administered in 22 languages, which had found 12.64% of contributors were female. (Glott & Ghosh, 2010). Here, there is no discipline-specific explanation for the observed gender gap.

Hemerly's survey of MusicBrainz adds to this data on the existence of gender gaps in online peer production communities. Hemerly's survey respondents were 97.18% male and editors interviewed suggested that very few women were active in the community. (Hemerly, 2011). The enjoyment of music and the desire to have accurate metadata about one's music collection is not something one would expect to have such a significant gender skew. With this additional data point it becomes more difficult to dismiss the Wikipedia results as aberrational. An as yet unexplained factor or factors at work in these communities appears to affect the number of female contributors. Further research on this issue is essential.

Age

FLOSS (Ghosh, 2005)	**Wikipedia (Glott & Ghosh, 2010)**	**MusicBrainz (Hemerly, 2011)**
10-21 (48.65%)	10-21 (48.7%)	18-21 (7.66%)
22-34 (45.88%)	22-29 (27.4%)	22-34 (63.31%)
35-44 (4.31%)	30-85 (24%)	35-44 (20.97%)
45-54 (0.99%)	[raw data requested	45-54 (6.45%)
55+ (0.17%)	from authors]	55+ (0.16%)

It would appear that MusicBrainz attracts a slightly older group of contributors than either free software projects or Wikipedia, but perhaps more importantly the majority of contributors to all three communities are under 35.

Family

The data on age makes it tempting to imagine a picture of the typical contributor as young, unmarried, and without child-care responsibilities. However, the FLOSS survey found 39.9% lived with a partner or spouse and that 17% reported having children. The Wikipedia survey also reported that, while women are a minority of the contributors, after the age of 32 has passed women in this cohort spend more than 2 hours more per week than men creating Wikipedia content. The authors note this as an area for further study and speculate that women at this age are less often full-time employed, often stay at home in order to care for children, and often work as freelancers, and that some of these factors may play a role here. (Glott & Ghosh, 2010).

Contributor-reported Motivations

Hemerly found that MusicBrainz contributors reported motivations that are often associated with the motivations discovered among free software and Wikipedia contributors, namely a philosophical preference that a commons be created and

maintained, a sense of belonging to a community, and sheer enjoyment of the activity. (Hemerly, 2011). Hemerly also found contributors frequently reporting a near compulsion to contribute out of a need for the metadata to be correct and complete. There are hints of this in other cultural commons and it deserves further study.

Philosophy

Contributors to free software projects are often characterized as sharing the view that software ought to be a commons, should be shared—in short—should be *free,* typically due to the perceived benefit to the greater good. Various studies have found developers stating that they "think that software should not be a proprietary good" (37.9% gave this as a reason for remaining active in a free software project, Ghosh et al., 2002) or that they "believe that source code should be open" (33.1% Lakhani & Wolf, 2005).

Surveys of Wikipedia editors have found similar motivations. The highest number of respondents, 72.91%, selected "I like the idea of sharing knowledge and want to contribute to it" as their motivation for contributing. (Glott et al, 2010). Similar statements fared well: "Because I like Wikipedia's philosophy of openness and collaboration." (30.07%) and "Because I think information should be freely available to everyone " (37.86%).

Hemerly found that MusicBrainz contributors overwhelmingly agreed (or strongly agreed) with statements such as "Metadata should be free and open to all." (97.88% of those responding) and "Information resources like MusicBrainz, peer-produced or otherwise, should be free." (98.70% of those responding). (Hemerly, 2011).

Community

Researchers have also found that free software developers and Wikipedia contributors cite a sense of belonging to a community as a key motivator. [cites]

Hemerly found that 88.60% of MusicBrainz contributors agreed or strongly agreed with the statement, "As a contributor, I feel part of a community and its mission." (Hemerly, 2011).

Enjoyment

As early as 1985, Richard Stallman explained in his GNU Manifesto that "Programming has an irresistible fascination for some people, usually the people who are best at it... creativity is also fun, a reward in itself." (Stallman, 1985). Researchers have found the same motivation from survey respondents. In one study, the top single reason to contribute to free software projects (44.9 percent)

was based on an enjoyment-related intrinsic motivation: "Project code is intellectually stimulating to write." (Lakhani & Wolf, 2005).

Many surveys of Wikipedia contributors fail even to ask whether contributors participate for fun, but in one survey that did inquire about such a motivation it was the highest-rated response. (Nov, 2007).

Hemerly also found MusicBrainz contributors overwhelmingly agreeing and strongly agreeing (90.91%) that "Contributing to MusicBrainz is fun" with contributors also describing it as intellectually stimulating, relaxing, and meditative. (Hemerly, 2011).

Compulsion

Hemerly also had several contributors volunteer another reason for their contributions, which they self-described as "OCD" or "obsessive compulsive" though without sincerely implying an actual clinical diagnosis. (Hemerly, 2011). In the case of Wikipedia, the second most-listed motivation in the Glott survey was "I saw an error I wanted to fix." at 68.78% (Glott et al, 2010), so this motivation may also be present in other online collaborations. It may also arise only when the task modularization has become as fine-grained as it has become in Wikipedia and MusicBrainz, where the ability to engage in a very small act that increases order becomes too tempting for some to resist.

Towards a Theory of Successful Online Peer-Produced Information Commons

Across free software projects, Wikipedia, and MusicBrainz, one can identify common features of the resources created by these communities, of the active participants in these communities, and the motivations that they report. In short, the features that Ostrom found important in natural resource contexts (Ostrom, 1998) can, with some adaptation, apply to these information commons as well.

When individuals are presented with an information commons with plausible promise and modular tasks, where they are given autonomous access to the resource and can predictably discern its current condition and availability, where it meets a repeated need of the individual, where individuals can participate in modifying the community's operational rules, and where conflict-resolution mechanisms are available, individuals may be motivated to contribute to such an information commons to the extent it presents an opportunity to join a community engaged in doing good and having fun.

References

- Hemerly, Jess (2011) *Making Metadata: The Case of MusicBrainz* , http://www.ischool.berkeley.edu/files/student_projects/jhemerly_mims_finalproject_0.pdf
- Glott, Ruediger and Ghost, Rishab (2010) *Analysis of Wikipedia Survey Data, Topic: Age and Gender Differences*, UNU-MERIT, http://www.wikipediastudy.org/docs/Wikipedia_Age_Gender_30March%202010-FINAL-3.pdf
- Ghosh, Rishab Aiyer (2005) *Understanding Free Software Developers: Findings from the FLOSS Study*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE, 23-45, The MIT Press.
- Gracenote (2010) *Gracenote | Global Media Database*, http://www.gracenote.com/products/global_media_database/
- Lakhani, Karim R. and Wolf, Robert G. (2005), *Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects*, in PERSPECTIVES ON FREE AND OPEN SOURCE SOFTWARE, 3-21, The MIT Press (2005).
- Lam, Shyong (Tony) K., Anuradha Uduwage , Zhenhua Dong , Shilad Sen, David R. Musicant , Loren Terveen , John Riedl (forthcoming) *WP:Clubhouse? An Exploration of Wikipedia's Gender Imbalance* , Proceedings of the International Symposium on Wikis and Open Collaboration, http://grouplens.org/system/files/wp-gender-wikisym2011.pdf
- Lerner, Josh, Tirole, Jean (2002) *Some Simple Economics of Open Source*, The Journal of Industrial Economics, Vol. L No. 2, 197-234.
- Margolis, Jane and Fisher, Allan (2002) UNLOCKING THE CLUBHOUSE: WOMEN IN COMPUTING, The MIT Press.
- MetaBrainz (n.d.) Finances, http://metabrainz.org/finances/
- MusicBrainz Blog (2011) *Our hardware fundraiser is complete: We raised $15,527.50*, http://blog.musicbrainz.org/?p=802
- Nov, Oded (2007) *What Motivates Wikipedians*, Communications of the ACM Vol. 50 No. 11, https://dl.acm.org/citation.cfm?id=1297798
- Ostrom, Elinor (1998) *Reformulating the Commons*, http://www.scielo.br/pdf/asoc/n10/16883.pdf
- Raymond, Eric Steven (1997) *The Cathedral and the Bazaar*, http://catb.org/~esr/writings/homesteading/cathedral-bazaar/
- Robles-Martínez, G., H. Scheider, I. Tretkowski, and N. Weber (2001) *WIDI: Who is doing it?*, Technical University of Berlin, http://widi.berlios.de/paper/study.html
- Stallman, Richard M. (1985) *The GNU Manifesto*, Dr. Dobb's Journal of Software Tools, Vol. 10, No. 3, http://www.gnu.org/gnu/manifesto.html

- Tapscott, Don, and Williams, Anthony D. (2006) WIKINOMICS: HOW MASS COLLABORATION CHANGES EVERYTHING, Penguin Group.
- von Krogh, Georg, Spaeth, Sebastian, Lakhani, Karim R. (2003) *Community, joining, and specialization in open source software innovation: a case study*, Research Policy 32, 1217-1241.