Editor: James Hendler
University of Maryland
hendler@cs.umd.edu

MusicBrainz: A Semantic Web Service

Aaron Swartz

Music has always caught the public's imagination. From dreams of a giant "jukebox in the sky" over the Information Superhighway[1] to the recent debate about Napster, music has always been the "killer app" used to describe new technologies. Of course, these dreams have never quite come about as planned. Instead of a smart machine seeking out music tuned to my tastes, I still have only a small number of choices on my radio dial. And ever since Napster started filtering, sharing music on the Internet has become increasingly difficult.

One thing that underlies these ideas is their dependency on *metadata*, or data about data. Metadata provides information about artists, song titles, and so on. All that information is attached to the music, but isn't part of it. The music world suffers from a lack of standardization in terms of metadata formats, as well as a paucity of public metadata.

The MusicBrainz project

The MusicBrainz project hopes to change this situation. It's a large database of music metadata, and even though it's only in beta testing right now, it already contains over 300,000 tracks. MusicBrainz information is all user-contributed, providing what some have termed the "cornucopia of the commons."[2] Unlike many situations, where each user decreases the value of the shared space (the so-called "tragedy of the commons"[3]), the easy duplication of electronic information creates a situation where each user makes the system more valuable.

For example, when you purchase a new CD and insert it into your computer, your audio player will probably come up with a generic name for it, *Audio CD 47* perhaps, complete with "Track 1," "Track 2," and so on. Were you to use MusicBrainz, the audio player would have attempted to connect to the MusicBrainz server, to see if that CD's metadata were available. If it were, your player would have renamed the CD (*Amnesiac*) and the tracks ("Pyramid Song" and so on). If metadata were not available, you likely would have filled in the track names yourself, if only for your own benefit. Your player would then ask you if you wanted to share

the information; if you did, it would send the information to MusicBrainz to share with other users. This type of functionality is currently implemented in an audio player called FreeAmp, and MusicBrainz plugins hopefully will be released for other players soon.

MusicBrainz origins

MusicBrainz wasn't the first to implement this idea. Back in 1996, the Internet Compact Disc Database (CDDB) was created using a similar system. It grew incredibly fast, with its users contributing track and title infomation for 800 new CDs each day. However, it had no moderation system and, thus, soon became filled with typos, misspellings, and duplicate data. Worse, Gracenote, a technology company specializing in content delivery, later bought it and imposed severe restrictions on its use. The many contributors who had helped populate the database were outraged that they could no longer use it freely.

Designed for the users

To replace CDDB, many projects sprang up. One of them, the CDIndex, later became MusicBrainz. To prevent a CDDB-style takeover, MusicBrainz implemented several safeguards. The database is freely available to the public under the OpenContent license, ensuring that no one can take control of the information without giving back to the community. Also, MusicBrainz set up a distributed server network, so that no one server has full control of the database.

Of course, MusicBrainz is adding features that weren't available with CDDB. Using audio fingerprinting software from Relatable, it can also get music metadata for an MP3 file. In addition, to prevent the problems that CDDB had, MusicBrainz includes a moderation system that lets users correct mistakes in the database.

Semantic Web services

MusicBrainz is one of the first of what might be called Semantic Web Services. These are a combination of the ideas behind the Semantic Web and Web Services. The Semantic Web aims to add machine-processable informa-

tion to the largely human-language content currently on the Web. Web Services is a similar concept aimed at sending machine-processable information between organizations in an attempt to automate processes.

MusicBrainz does both. In addition to providing its content in HTML, MusicBrainz also provides its information in Resource Description Framework, the Semantic Web's foundational language. It gives all the major items in its database (artists, albums, tracks, and so on) URIs, so that others can refer to them and use them in numerous applications. All of this is done using open protocols, so that anyone can easily access the data for any purpose. MusicBrainz also provides an RDF-based application programming interface so applications can query the database and submit new information. MusicBrainz also provides an open-source library so programmers can easily contact the MusicBrainz API and include the added functionality in their applications.

Because the MusicBrainz format is open and uses RDF, the industry standard for metadata, it can be repurposed for numerous applications. For example, other people can build on MusicBrainz' data. Music vendors can use the MusicBrainz metadata to enhance the information on their Web sites and add information about where to legally purchase the music.

File-sharing systems (such as Napster, Freenet, or Audio Galaxy) can use the metadata to provide more information about the MP3s that are available for download or to make it easier to search for a song. Artists can provide links to a "tip jar" in which appreciative fans can donate money if they like the music. In fact, the Espra project has built a decentralized architecture on top of Freenet and music metadata to provide a direct link between artists and their fans. This attempt to build a gift economy is similar to the patronage system, which funded some of the greatest artists in history.

Other projects, which might not want to use the MusicBrainz database, can still use the RDF standards that it has defined (such as "artist," "album," and so on) for their own databases. This provides them with a "semantic bootstrap"—access to terms that already have a well-defined meaning and are widely used.

Because the RDF format is so flexible, adding data will not affect backward compatibility with clients that don't expect it or use it. Because these are all open protocols,
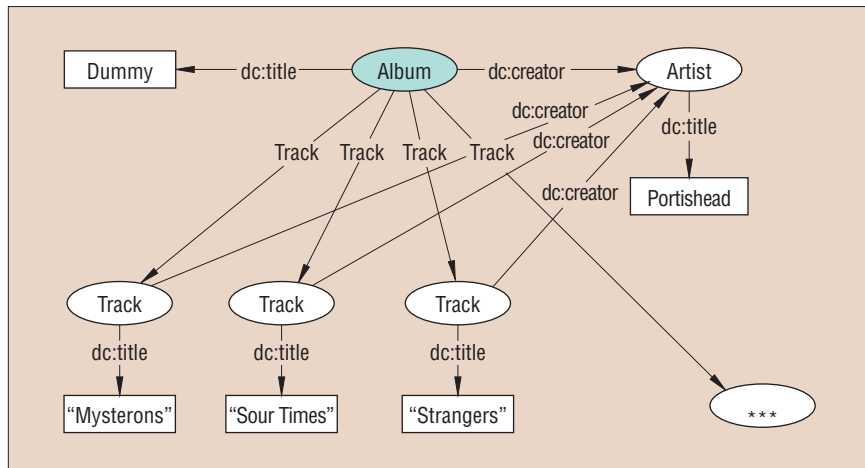

Figure 1. A graphical representation of MusicBrainz RDF data about a Portishead album.

More on MusicBrainz

Robert Kaye, MusicBrainz's Mayhem and Chaos Coordinator, led the development of MusicBrainz (www.musicbrainz.org) using international programmers. The source code is available as free software. The project is sponsored in part by Bitzi, Relatable, EMusic, O'Reilly and Associates, and its users.

FreeAmp is a free audio music player that uses the MusicBrainz database. Grab a copy at http://freeamp.org.

Espra is an open-source file-sharing client that attempts to produce a decentralized system for the distribution of information and the creation of a gift economy. Go to www.espra.net for more information and a free download.

clients have a choice of which server they want to receive their data from, and some might want to choose servers that provide them with higher-quality (or more) information. All the servers are compatible, so users can easily switch to another.

Next stop: The Semantic Web

Semantic Web Services such as MusicBrainz are an exciting part of the growing number of Semantic Web tools and applications. The key to the Semantic Web's power is in building apps like these that publish information. Once such systems begin talking with each other, the possibilities are endless. Their combinations can reveal potential that we wouldn't have thought of before, and empower people—such as music fans long separated from artists by the record companies—to take control of their world. It's an exciting ride! ■

References

1. M.L. Detourzos, *What Will Be: How the New World of Information Will Change Our Lives*, Harper, San Francisco, 1997.

2. D. Bricklin, "The Cornucopia of the Commons: How to get Volunteer Labor," www.bricklin.com/cornucopia.htm (current Jan. 2002).

3. G. Hardin, "The Tragedy of the Commons," *Science*, vol. 162, no. 3859, Dec. 1968, pp. 1243–1248; http://dieoff.com/page95.htm.
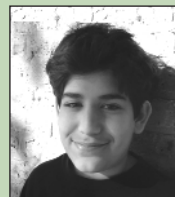
Aaron Swartz is a contributor to the MusicBrainz project, especially in its metadata initiative. He is a member of the W3C's RDF Core Working Group and a cofounder of the Semantic Web Agreement Group (SWAG). His latest project is the Plex (plexnow.com), a decentralized storage and query system for RDF data. Contact him at me@aaronsw.com or through www.aaronsw.com.