

Media Streams Metadata Exchange User Study and Initial Requirements

Ryan Shaw

MIMS Final Project Report

May 20th, 2005

Table of Contents

Executive summary.....	1
Overview.....	2
Meeting the need for semantic metadata.....	2
Related work.....	2
Semantic annotation systems.....	2
Commons-based peer production.....	3
User-generated metadata.....	4
User study: the <i>anime</i> music video community.....	5
Data gathering.....	5
Interviews.....	5
Online observation.....	6
Survey	7
Findings.....	8
Experience model.....	8
Personas.....	9
"As-is" scenarios.....	11
Scenario 1A: The remixer.....	11
Scenario 2A: The enthusiast.....	14
Scenario 3A: The newbie.....	14
Collaboration model.....	14
Proposed mediating technologies.....	15
Media-on-demand network.....	16
Shot metadata repository.....	18
Remixing engine.....	19
"To-be" scenarios.....	20
Scenario 1B: The remixer.....	20
Scenario 2B: The enthusiast.....	21
Scenario 3B: The newbie.....	22
Matching shot requests to shot descriptions.....	22
Request analysis.....	23
Annotation experiment.....	25
Comparison.....	27
Conclusion.....	28
Initial MSMDX platform requirements.....	29
User requirements.....	29
Use cases.....	30
Functionality requirements.....	32
Public API.....	32
Media player plugins.....	33
Structured metadata plugins	33
Sequencing tool.....	34
Editor plugins.....	34

Executive summary

In recent years technologies for producing multimedia content have become cheap and ubiquitous, while the effective bandwidth for distributing this content has increased, due to greater network capacity and new protocols and formats for using this capacity more efficiently. Unfortunately, the development of applications for finding, sharing, and recombining this content, rather than simply consuming it, has not kept pace. The MSMDX (Media Streams Metadata Exchange) project plans to remedy this by developing a platform to enable users to annotate, find, share, and remix audio and video content in a networked, collaborative fashion. This spring a single community of active audiovisual content consumer/producers was studied in order to understand their current practices and how they might be supported or changed by new tools. The results of this study have resulted in a set of initial user requirements for the MSMDX platform. This report describes the study in detail and presents the requirements for three components of the platform:

- A system for providing fragments of media on demand
- An open repository for user-contributed descriptions of media
- An engine for applying user-contributed rules to generate remixed media

Overview

Meeting the need for semantic metadata

The rise of broadband networking, peer-to-peer protocols for sharing large files over the Internet, massive drops in the price of digital storage, advanced techniques for compressing audio and video, and the widespread distribution of cheap devices for producing and consuming media have resulted in an explosion in the amount of multimedia information (images, video, and audio) available on the Internet. Yet finding specific content of interest in this mass of multimedia is still difficult, and while it has become fairly simple to upload raw audio or video onto the Web, fluidly combining multimedia materials in the way we have become accustomed to doing with text is still a dream.

The key is to provide metadata—descriptions of the content and structure of multimedia information—that enable humans and computers to find, use, share, and remix networked media according to descriptions of their semantic content and rules for recombining them. But creating the necessary metadata is not easy. Attempts to create algorithms that can do the job automatically have largely failed, unable to make the leap from the low-level features perceptible by machines to the higher-level features desired by human users. Manual approaches to creating higher-level metadata—paying people to annotate media after it is created—are labor-intensive and expensive.

The MSMDX project reconceptualizes the problem of how to generate semantic metadata for multimedia as a question of how to design a socio-technical system that encourages large communities of peers to produce metadata for free, as a byproduct of their interactions with media. Rather than paying a small number of experts to annotate media, tools could be designed in such a way as to allow anyone who is interested to create annotations, which could then be used by anyone else. There is evidence that this kind of open, distributed system can scale to the level needed for the production of metadata for the mass of multimedia content on the Web.

Since the proposed solution involves not just the development of new technologies, but the development of a system composed of both technologies and cooperative networks of humans, its design must be informed by a solid understanding of the social context in which it will be deployed. In particular, we must try to understand how participants in communities formed around media content interact with each other and with media, and how these interactions may change as new technologies are introduced. This report summarizes an investigation of one such community and the socio-technical design requirements that were uncovered.

Related work

The ideas which form the foundation of the MSMDX project have their roots in a number of areas, including research in the multimedia community to develop systems for the creation of semantic metadata, projects which harness the collective production capacity of large groups using cheap and ubiquitous communications technologies, and efforts to open the creation of metadata for information to the users of that information.

Semantic annotation systems

There have been a number of systems developed in the multimedia research community to enable the manual creation of semantic metadata for multimedia content. In particular, the Media Streams system allows users to represent a number of different aspects of video content including *mise-en-scene*, characters

and objects, actions, and cinematography.¹ Users can take advantage of these representations to query for content of interest, and to generate new works by recombining them.

However, most of these annotation systems, like Media Streams, focus on individual users, not cooperative networks of users. Those projects that do enable cooperative work have generally focused on professional users such as media professionals, film scholars, and educators. This focus limits the scale of metadata production by raising barriers to contribution, and creates a situation in which the metadata produced is considered a strategic asset and cannot be shared freely.

Fortunately, there are models for cooperative activity around a freely shared collection of information which show how systems which transcend these issues might be designed.

Commons-based peer production

In recent years a number of successful projects have emerged which rely on the cooperative activity of large groups of volunteers. Two of the most well-known projects are Linux, an open-source operating system developed by a distributed coalition of programmers from around the globe,² and Wikipedia, a high-quality online encyclopedia maintained through the collective efforts of thousands of contributors.³ These projects show that it is possible to build high-quality information products by facilitating large-scale cooperation among unpaid volunteers.

Law professor Yochai Benkler has named this phenomenon “commons-based peer production,” and he suggests that it “has systematic advantages over markets and managerial hierarchies when the object of production is information or culture, and where the physical capital necessary for that production—computers and communications capabilities—is widely distributed instead of concentrated.”⁴

Benkler argues that successful peer production systems exhibit three characteristics. First, they must be modular, meaning the work to be done can split into pieces and handled incrementally and asynchronously. The creation of semantic metadata for multimedia fulfills this criteria—metadata can be created for a given piece of media in stages, by different people working independently. Second, the modules of work must be highly granular, meaning that contributors can make even very small contributions to the overall effort. Lack of this granularity has been a problem for professionally-focused multimedia annotation systems, which generally are too complex to allow lightweight contributions.

Finally, successful peer production systems must have automated or semi-automated systems for integrating contributions and exercising quality control. These systems may be under centralized control, as in the case of Linux, or control may also be distributed to the community, as in the case of Wikipedia. A successful system for the peer production of multimedia metadata will likewise need to be designed in order to maximize the quality of contributions, by providing incentives to contribute quality information and allowing poor-quality information to be identified and weeded out.

1 Marc Davis, "Media Streams: An Iconic Visual Language for Video Representation," in *Readings in Human-Computer Interaction: Toward the Year 2000*, 2nd ed., ed. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg, 854-66 (San Francisco: Morgan Kaufmann, 1995).

2 Linux Online, "What is Linux," <http://www.linux.org/info/index.html>.

3 *Wikipedia*, s.v. "Wikipedia," <http://en.wikipedia.org/wiki/Wikipedia>.

4 Yochai Benkler, "Coase's Penguin, or Linux and The Nature of the Firm," (2002), <http://www.benkler.org/CoasesPenguin.html>.

User-generated metadata

A parallel phenomenon that has gained prominence recently is the rise of user-generated metadata. Traditionally metadata has been created by professionals such as librarians. However, professional creation of metadata has proven unable to scale to the massive amounts of information being created and made available through the Web. An alternative approach is to have creators of information add their own metadata. This is more scalable, but suffers from the problem that creators of information do not have a good understanding of how users might conceptualize their work. Worse, unethical creators may purposefully mislabel their works in order to trick users into buying their wares.

A third possibility is to allow not just creators, but users of information to create metadata for that information. This is the approach taken by a number of popular web sites including the del.icio.us social bookmarking service,⁵ and the Flickr photo-sharing service.⁶ By allowing users to create metadata that is useful for them personally, aggregating large amounts of this personal metadata, and making it publicly available, these services have succeeded in both encouraging users to create metadata and making that metadata useful to others. These sorts of community-built and -maintained metadata frameworks have been dubbed “folksonomies.”⁷

However, folksonomies as currently implemented have a number of drawbacks. Since they rely on strings or “tags” to label content, without any framework for structuring these tags, there are problems with both polysemy (tags which have multiple meanings, like “cal” being used both as a nickname for UC Berkeley and as a person's name) and synonyms (different tags being used to mean the same thing, like “movies” and “cinema”).

Some folksonomy proponents argue that these problems are actually features, and that fine distinctions in meaning between terms like “movies” and “cinema” will be lost if synonym control is introduced.⁸ This is a good point, and it suggests a need for a framework which can allow some users to maintain these distinctions while others, less concerned with such subtleties, can choose to collapse them. This could be achieved through the introduction of another level of community production in which users produce not only individual tags, but also conceptual links among those tags.

5 Joshua Schachter, “About del.icio.us,” <http://del.icio.us/doc/about>.

6 Flickr, “About Flickr,” <http://flickr.com/about.gne>.

7 Adam Mathes, “Folksonomies - Cooperative Classification and Communication Through Shared Metadata,” (2004), <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.

8 Clay Shirky, “Folksonomy,” (2004), <http://www.corante.com/many/archives/2004/08/25/folksonomy.php>.

User study: the anime music video community

The failure of previous multimedia metadata systems to scale to the Web and the success of commons-based, user-contributed approaches to building information systems indicate a need for research into a community platform for the production of multimedia metadata. Since it is meant to mediate between participants in a community, it is important that the technical design of the platform be informed by an understanding of actual social practices and processes. Toward that end, a community of active audiovisual content consumers and producers was studied in order to understand their current practices and how they might be supported or changed by new tools.

The study had several goals:

- To catalyze creative ideas for innovative tools
- To develop “real world” use cases for prioritizing design ideas
- To develop a model of what users of the tools might be like, including demographics, technologies currently used, existing practices, and consumption/production patterns

The community in question is centered around the creation, sharing, and discussion of *anime* music videos (AMVs). AMVs are fan-produced music videos which incorporate content from *anime* (Japanese animation) films and television series, setting them to popular music. The creation of music videos by fans of popular media content is not new; in fact the practice has been widespread since the introduction of videocassette recorders to the general public in the late 1970s. But the AMV community has taken the practice to new levels of sophistication, using prosumer-level non-linear video editing tools to create incredibly professional productions, which are then shared over the Internet and shown at AMV contests at *anime* fan conventions around the world.

The study resulted in a set of personas representing the different roles played by participants in this community, as well as detailed scenarios describing their current practices, and how these practices might change given new tools for sharing metadata. The study found that there interest in tools sharing metadata and that such tools could potentially bridge the gaps that currently exist between remixers and consumers in this community's media and metadata economy. But in order to find acceptance, these tools must be designed to augment rather than replace existing practices.

Data gathering

A combination of qualitative and quantitative methods were used to gather data on the AMV community. First in-depth interviews were conducted with remixers (creators of AMVs), and online discussions, FAQs, wikis, and other public artifacts produced by the community were observed and analyzed. The context and background provided by these investigations then informed the design of an online survey.

Interviews

Three remixers were interviewed. Initially more interviews were planned, but recruitment turned out to be a problem. Although the AMV community is quite large, it is also widely distributed. All the interviewees remarked on the fact that they knew very few, if any, people in their geographical area who participated in the community. Most face-to-face interaction occurs at the fan conventions held at regular intervals throughout the year. Efforts were made to recruit more interviewees at one such convention which occurred this spring in San Francisco, but it unfortunately turned out to have only a very small AMV

contingent.

The first interviewees were two males, ages 21 and 24. They were both members of the UC Berkeley *anime* club and friends with one another. Both had been fans of *anime* for about seven years, and had been involved in the AMV community for about four years. Contact was established by using the *anime* club's web site to contact the officers of the club, who suggested that these two remixers. They were interviewed as a pair, and the interview was video recorded.

This was an extremely fruitful interview, as the pair format enabled the two interviewees to elicit comments from one another that probably would not have emerged in a solo format. One interviewee was a novice, having only created one music video, while the other was fairly experienced. They also differed in their creative approaches, with one favoring a more “purist” approach of creating music videos solely through temporal editing, and the other experimenting heavily with sophisticated effects like rotoscoping and layers. These differences and the discussions they catalyzed proved to be very informative.

The third interviewee was a 24 year old male, contacted at the aforementioned fan convention. He had been a fan of *anime* for about ten years and involved in the AMV community for two years. In terms of creative experience he fell somewhere between the previous two interviewees, with four music videos under his belt. Since this interviewee works for a software company which sells video editing software, he was able to give a detailed explanation of some of the technical issues surrounding music video editing.

All the interviewees were very reflective about their practice. It was clear that these remixers took their work seriously, and they appreciated the serious tone taken while interviewing them. All the interviewees expressed great interest in the project and a desire to stay involved.

Since all the interviewees were males in their early 20s, there was initially some concern that the group of interviewees was not diverse enough. Later it was learned that the group was representative of the demographics of the larger community.⁹

Online observation

The AMV community, while quite distributed in the physical world, is rather concentrated in the online world, gathering mostly around a single website, AnimeMusicVideos.Org (henceforth AMV.Org).¹⁰ This turned out to be rather fortunate, as the study could focus on a single (albeit large) site. The purpose of the site is stated in the site FAQ:¹¹

This web site is to create a place where people who enjoy and create anime music videos can get together, share ideas, learn from one another, and ultimately have everyone creating and enjoying better videos.

The ultimate goal of the website is to make a database of every Anime Music Video ever made.

In addition to hosting over 59,000 AMVs, AMV.Org also provides “how-to” guides and tutorials on creating videos, guides to hardware and software for creating videos, interviews with remixers, online journals for site members, chat rooms, and web-based bulletin board-style forums for discussing *anime* and AMVs. While some of this is available to the public, most of the content and features require free registration, and the site has over 300,000 registered members. AMV.Org does not run any advertising, and

9 103 of 125 (82%) of survey respondents were male and their median age was 20 years old. Furthermore, a number of posts to online forums to the effect of, “Why don't more girls do this?” were observed.

10 Online at <http://www.animemusicvideos.org/> or <http://a-m-v.org/>.

11 AnimeMusicVideos.Org, “Site FAQ: Website, Membership and Downloads,” <http://www.animemusicvideos.org/help/>.

relies on member donations to keep running. Donating to the site gives members access to advanced search functionality and accelerated downloads of videos.

The FAQs, how-to guides, and other background materials on the site provided a good overview of how AMVs were created, as well as an understanding of the community's "rules," such as the prohibition on trading digital files of original *anime* content. This was very useful for designing the survey.

Of particular interest was the "AMV Suggestions" discussion forum, used for posting ideas about AMVs that one would like to see created, as well as requests from remixer for pointers to appropriate *anime* shots and songs for their videos. As much of this discussion involved queries for and descriptions of media content, it was directly relevant to the goals of the study.

Survey

Once a preliminary understanding of the AMV community was formed through interviews and online observations, a 40-question survey was designed to obtain a broader picture of its demographics, how its members find content of interest, how they use that content, how they communicate and share both the original content and new derivative works they have created, and the current technologies they use for these various activities.

The survey was organized into five sections, with questions about *anime* consumption, AMV consumption, AMV creation, AMV creation technologies, and general demographics respectively. An open source surveying software package¹² was used to create the survey, which allowed the conditional presentation of questions based on previous answers. This was quite beneficial as it enabled respondents to avoid seeing irrelevant questions.

The survey was pre-tested on the three interviewees. All three received via email invitations to take the survey. Shortly after each person had taken the survey, he was phoned and asked for his reactions, what questions he found confusing or poorly worded, and suggestions for improvement. After being asked for general comments the pre-testers were asked for specific comments on each question in the survey. One person was unable to speak on the phone and emailed detailed comments instead.

Because many of the questions in the survey were based on information gathered in the interviews, the pre-testers generally found the questions relevant and intelligible. Thus most of the direct suggestions for improvement were minor but important fixes to the existing questions. The pre-test also indirectly resulted in some new questions being added, as seeing actual answers to the questions prompted the realization that some questions had gone unanswered. In particular, a question about reactions to proposed ideas for new technologies was added in order to get a sense of what ideas most interested the respondents.

Initially the survey was planned to be administered to a random sample of the AMV.Org membership, but the administrators of the site did not feel comfortable allowing this. They suggested a link to the survey be posted on the AMV.Org "General AMV" forum. This was not ideal, as it meant that respondents "opted in" to taking the survey themselves, rather than being selected. "Opt-in" surveys are subject to self-selection bias; in other words, the sample may not reflect the community as a whole, but the community of people who want to be surveyed. This bias may have been further exacerbated by the formal tone of the invitation—most of the forums discussions are quite informal.

125 people responded to the survey, which ran for a period of two weeks in mid-March. Although this is a respectable number, it is probably too small for the size of the community. Furthermore, 109 (87%) of

12 PHP Surveyor, <http://phpsurveyor.sourceforge.net/>.

the respondents claimed to be remixers, which does not reflect the overall site statistics of 20,518 remixers among 329,839 members.¹³ This may be a result of the self-selection bias mentioned above, or it may reflect the fact that mainly remixers frequent the “General AMV” forum. Finally, 57 of the respondents provided contact information and indicated that they would be interested in participating in future work.

Findings

The data gathered from the survey and interviews were analyzed to produce a portrait of the AMV community, including a model of how participants' roles changed with experience, a set of personas representing these various roles, scenarios of representative interactions among members of the community, and a framework for understanding the cooperative links between these members.

Experience model

Participants in the AMV community have varying degrees of technical and artistic skill and varying amounts of knowledge about *anime*. These skill and knowledge levels vary with experience. It is important to note that skill at creating AMVs is orthogonal to knowledge about *anime*. One may be a skilled remixer but have limited knowledge about the wide range of *anime* characters and narratives. Or (as is more common in this community) one may have extensive knowledge of *anime* but lack the skill or access to equipment needed to create AMVs.

As participants advance among one or both of these axes, the nature of their participation changes. John Fiske's categories of cultural production¹⁴ are useful for analyzing these different patterns of participation. Fiske understands popular culture not as music, movies, and TV shows, but as something that is produced by people out of those cultural products. He categories this cultural production into three areas: semiotic production, enunciative production, and textual production.

Semiotic production, the creation of personal meaning out of the media we see and hear, is something that any person who interacts with cultural products engages in, not just fans. Though movie and television viewers, music listeners and video gamer players are often characterized as “consumers,” it is important to realize that they are actively creating meanings from the the products they consume, although they may not articulate these meanings to others.

Fans, on the other hand, are distinguished by the fact that they do articulate these meanings to others. This is enunciative production: gathering with other fans to talk about music or TV shows, discussing content of interest in online forums or on mailing lists, compiling episode guides and character databases. Engaging in enunciative production increases one's knowledge of cultural content, and as one's knowledge grows this enunciative production becomes more pleasurable.

Finally there is the form of production engaged in by AMV editors: textual production. Textual producers use cultural products as raw materials to create new cultural products, often with production values equal to or surpassing the originals. Textual production requires technical and artistic competence, which come with experience. As one remixer put it, “[S]ome part of the experience is also making crap along the way so you learn from your mistakes. I mean it's very surprising when someone on their first try makes something good.”

¹³ Site statistics as of May 11th, 2005.

¹⁴ John Fiske, “The Cultural Economy of Fandom,” in *The Adoring Audience: Fan Culture and Popular Media*, ed. Lisa A. Lewis, 37-42 (New York: Routledge, 1992).

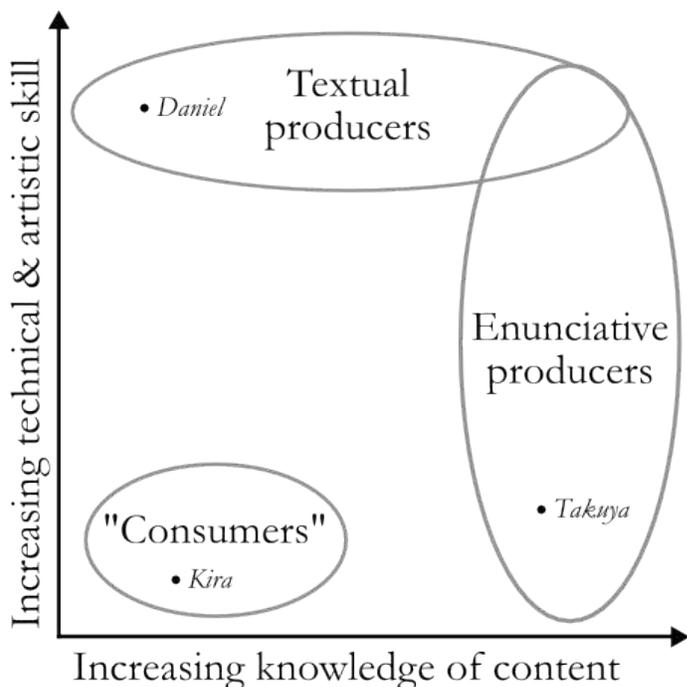


Figure 1. An experience model of the AMV community.

We can use these two orthogonal axes of development and Fiske's categories of production to build an experience model for the AMV community (Figure 1), showing how modes of participation can evolve over time. In the next section three composite personas will be presented that represent participants at different points in this model, represented in the figure below by their (fictional) names.

Personas

The following composite personas are based on the demographic information obtained from the survey, descriptions of roles in the community given by interviewees, and observation of the AMV.Org discussion forums. The quotes that introduce each persona are actual quotes from interviews or forum posts.

The personas are presented in order from the top to the bottom of the AMV community hierarchy. At the top are the skilled AMV editors, in the middle are knowledgeable fans, and at the bottom are the "newbies." These personas should not be interpreted as clearly distinguished categories: many editors are also knowledgeable fans, and roles can change over time: one of the editors interviewed admitted to having been "one of those people who always ranted about good ideas for music videos" before learning to edit himself.



Daniel Railer

Role: Remixer.

Male, 21 years old.

"I like to have something brand new every single video, but sometimes you can't make the whole thing brand new, different from everything else you've seen."

Background

Daniel is a senior at UC Davis, where he is majoring in biochemistry. (He tried to switch to

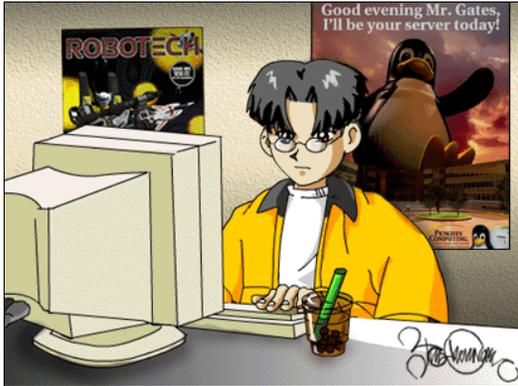
film studies at sophomore year but his dad threatened to stop paying for college.)

He's been making AMVs for three years, and has made 11 videos. He started out using Windows Movie Maker but quickly moved up to Adobe Premiere. Now he's starting to learn Adobe After Effects so he can do some more complicated compositing.

Daniel doesn't watch as much *anime* as he used to, but he still loves making music videos. He's won a few contests at local fan conventions, but nothing big yet. He likes the competitions because they motivate him to finish his videos, but he doesn't think much of the judging. In his opinion, videos with cool bands that pander to young fans often win over original artistic visions. He hopes to change that with his current project.

Goals

- Find the right shots to finish his *magnum opus*
- Win "Most Original Concept" in the 2005 AMV.Org Viewer's Choice Awards
- Make finals at the Anime Expo 2005 AMV contest



Takuya Suzuki

Role: Enthusiast.

Male, 19 years old.

"I'm a trivia freak, so it just completely makes me ecstatic if I know something somebody else doesn't."

Background

Takuya is a sophomore at Grinnell College, where he is majoring in computer science (though he spends more time on his duties as an officer of

the Grinnell Anime Club). He's been a huge *anime* fan for about 8 years, and has seen hundreds of movies and TV series, including many that haven't been released in the United States.

He spends a lot of time on the AMV.Org discussion forums, especially the AMV Suggestions forum, where he uses his encyclopedic knowledge of *anime* to help music video editors find shots they're looking for. He is also an administrator at NarutoManiac, a website for fans of his favorite *anime* series, *Naruto*, where he helps keep the character biographies up to date.

Takuya doesn't make AMVs himself (he tried once but gave up after a few hours), but he is full of ideas. He often posts detailed scripts and storyboards for proposed music videos to the forums, but unfortunately, none of the editors seem to want to use his ideas.

Goals

- Earn respect for his *anime* knowledge
- See his visions for great music videos realized



Kira Posner

Role: Newbie.

Female, 14 years old.

"Hi, I'm new here and I've downloaded a lot of AMV's in my time here. Unfortunately, I am unable to make any videos as of yet; but, I have an idea for one... I really really really want to see this one done! Please, please!! "

Background

Kira is a freshman at Carter High. She's been into *anime* for a little over a year. When she gets home from school in the afternoons she heads straight to her computer to check for new music videos uploaded to AMV.Org.

She loves the music videos on the Org but wishes there were more videos made using the music of her favorite band, Black Remembrance. There are a few, but they use shots from Dragonball Z, which she hates.

Kira is a frequent poster on the AMV Suggestions forum, where she asks people to make cool videos for her. But she's getting tired of people telling her to "make her own." Her computer is too old to edit video on, and it doesn't even have a DVD drive! Plus she doesn't want to have to read a bunch of boring how-to guides.

Goals

- See more cool videos with her favorite bands and *anime*

"As-is" scenarios

The scenarios in this section describe some typical processes and interactions as they currently occur in the AMV community.

Scenario 1A: The remixer

This scenario is quite long, as it describes in detail the process of creating an AMV from the moment of inspiration. While the creative process is different for different remixers, the process described here is representative of standard practice, based on the information gathered from the survey and personal interviews. In particular, the importance of beginning with a song (rather than specific anime or an abstract idea) was emphasized by all the interviewees. The survey results supported this, with 46% of respondents beginning with a song, twice as many as the number who begin with an abstract theme, idea, or message (the second most common answer). In this scenario, Daniel begins with a song which develops into a more abstract theme involving unhappy students. Finally, the paper planning Daniel does was the most common technique cited by the survey respondents, about half of whom said that they map out or storyboard their music videos before beginning to edit.

Daniel is hanging out with his friend Kevin in Kevin's dorm room, listening to music on the computer. At one point a song comes on with some lyrics that catch Daniel's ear:

*I wonder why we waste our lives here
When we could run away to paradise
But I am held in some invisible vise*

Perhaps it's senior-year burnout, but the lyrics strike a chord with Daniel. The song reminds him of a series he saw at a recent Davis Anime Club meeting called *Cromartie High School*, about a kid stuck in a school full of tough delinquents. He thinks that maybe he could make a decent music video with this song. Since Anime Expo is coming up in a couple of months and he hasn't had any other good ideas yet, he asks Kevin to send him the MP3 of the song, which turns out to be Wang Chung's "To Live and Die in L.A."

When Daniel goes home that evening the song is in his inbox. Daniel does a web search for "live and die in l.a. lyrics." The first few results are for a 2Pac song by the same name, but the fourth result is the Wang Chung lyrics he's looking for. Reading through the full set of lyrics convinces him that this is a good candidate for an AMV. The lyrics paint a picture of a person trapped in a dangerous place, a perfect fit for the protagonist of *Cromartie High*. He prints out the lyrics and leaves them on desk, and goes to sleep that night listening to the song on his headphones, trying to memorize all the words.

Later that week, Daniel borrows the *Cromartie High* DVDs from the Anime Club. When he gets back to his room, he sits down at his computer, puts the "To Live and Die in L.A." MP3 on repeat play, and loads the first DVD into his DVD drive. With the music and *anime* playing at the same time he can get a good sense of how well they fit together. After watching the first DVD, he decides that there is probably enough footage to support the whole song, so he starts copying the DVDs to his hard drive.

Copying DVDs takes a while, so Daniel decides to do a little planning while he's waiting. He gets out a notebook and writes down all the lyrics to the song, leaving space between each line. He restarts the song in his MP3 player and watches the time elapsed display. At the start of each line in the song he pauses the player and writes down the elapsed time before each lyric. Finally he writes down notes about his ideas for shots to use in the space between lyrics (see Figure 2).

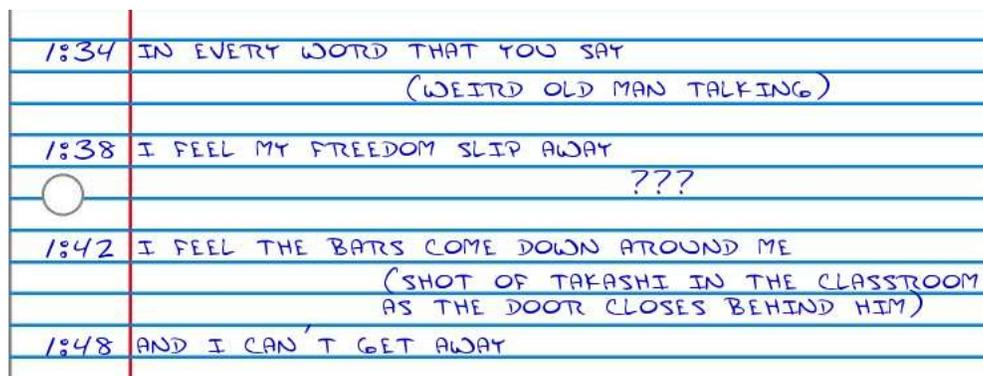


Figure 2. Part of Daniel's plan for his music video.

Finally the DVDs finish copying to his hard drive. Fortunately he has a very large drive, as the DVDs he's just ripped take up 20GB of hard disk space! He knows that editing the ripped video directly will be too much for his computer's processor to handle, plus he wants to free up some hard drive space, so he runs a program re-encode the ripped video into something simpler and lower quality.

At this point Daniel puts the project on hold for a little while, as the next step is horrifyingly boring. But a few weeks later he finds himself sitting at his desk with a midterm to study for. He decides that this is a good time to get some work done on his music video.¹⁵ He starts his video editing software and loads the first 90 minutes of video. Then he advances through the video, shot by shot, identifying shots that he wants to use and cutting them out into clips. He selects only the video track for each clip, as he isn't going to use the original audio. This is very tedious work, as he has to carefully adjust the in and out points to the beginning and ending frames he wants for each clip. But it's better than studying for a midterm. After several hours, Daniel finally has the clips he thinks he wants to use selected. He saves his project and goes to sleep.

Another few weeks pass before Daniel gets back to his project. Now he's excited, because he can actually start doing some creative work—everything up to this point has been boring setup work. He starts his editor and opens the MP3 file he first heard so many weeks ago. The song appears as a long waveform in his editor. Daniel consults the plan he has written out on a piece of notebook paper, and starts dropping video clips into place along the soundtrack. Soon he has the basic skeleton of the music video in place, although there are plenty of gaps to be filled in (see Figure 3).

Unfortunately, Daniel now runs into trouble. He realizes that some of the shots that he described in his paper plan don't actually exist in the footage he has. Other shots are there, but are too short for what he had planned to do with them. He reviews his clips several times and eventually comes to the conclusion that although he has the key shots, there isn't enough appropriate footage to complete his video.

But Daniel isn't ready to give up, having already invested this much time into his video. He still thinks his idea of interpreting the lyrics as the laments of an unhappy student is a good one. He just needs to find some more footage to fit that idea. There aren't any more episodes of *Cromartie High* he can use, and he can't think of any other suitable *anime* he could use. Finally Daniel decides to login to the AMV.Org "AMV Suggestions" forum, where he posts a message asking for help:

¹⁵ As one interviewee put it, "The key to making progress on *anime* music videos is to have something else to avoid."

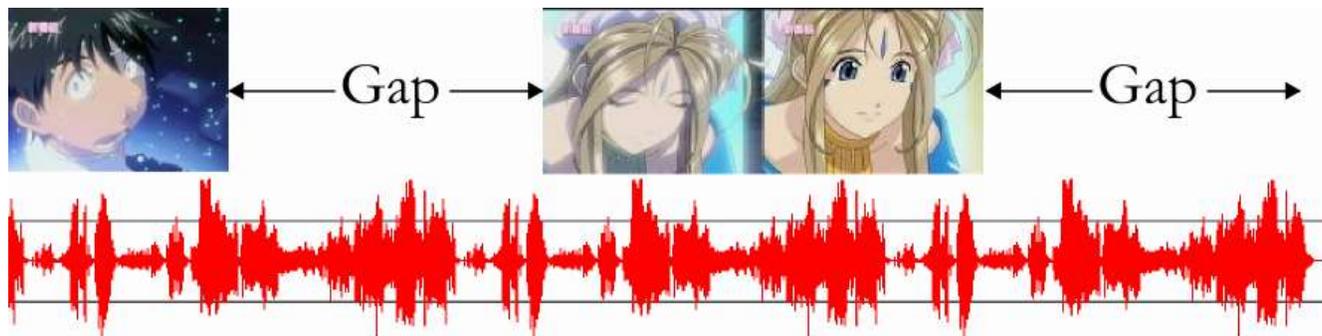


Figure 3. Gaps remain to be filled with appropriate shots.

I need some shots of students having a bad time at school. Shots of menacing bullies or prison-like school rooms would be especially good. I already have shots from Cromartie High School—any other suggestions?

A few hours later, Daniel gets an email notifying him that there has been a reply to his post, from takkun77. takkun77 writes:

What about Great Teacher Onizuka?? That has a TON of shots like that. For bullies, try the first season, second episode, when Onizuka faces off with Shinjiro's gang.

Daniel isn't familiar with *Great Teacher Onizuka*, but a few web searches confirm that it might be a good candidate. He logs on to his online DVD rental service and adds the complete run of *Great Teacher Onizuka* DVDs to his queue. A few days later they arrive in the mail. Daniel starts with the second episode, in order to check out the bully shot takkun77 mentioned. Sure enough, it's a good fit. He isn't looking forward to another round of ripping DVDs and pulling clips, but at least Daniel's music video is back on track.

Scenario 2A: The enthusiast

This scenario describes an expert fan who posts a detailed treatment of an idea for a music video to a discussion forum, a common occurrence based on observations of the AMV.Org forums.

It's Friday evening and once again Takuya is on the AMV.Org discussion forums. He's just finished another post to the "AMV Suggestions" forum about an idea he has for a music video:

Someone should really make a video using the song "Goodbye" by Disappearance with shots from Iron Intellectrix. If it were me, I'd use shots of Tran and Shiv trying to rescue their father, and of Eve and Nena. Shots from later in the series with the whole Eve "situation" would be great for the words "I'm your brain giving you someone to look at", and of course the "Has no one told you he's not sleeping" can go with a shot of Tran/Shiv digging up the grave or doing the transmutation. I'd imagine this being sort of in order, if you will. That is, not having the clips jump around, but build up to the end of the song in the order the events actually happened in the show.

He doesn't know why he bothers, though. Even when he posts detailed storyboards for music videos, specifying precisely which shots should be used with which lyrics, no one ever uses his ideas. And it's really a shame, too—*Iron Intellectrix* is a criminally overlooked series, and the Tran/Shiv narrative is begging for the AMV treatment. He wishes he could make his own video, but he just doesn't have the patience for all that fiddling with trimming frames and such. He'd rather leave that to the AMV editors.

It's really not fair, though, he thinks. He spends a lot of time helping AMV editors complete their visions—for example, earlier this evening he helped somebody find some shots of characters having a rough time at school. *Great Teacher Onizuka*, of course. The guy hadn't even heard of it before! You'd think that one of

the editors could return the favor by making a video based on Takuya's ideas, but no. Oh well...

Scenario 3A: The newbie

This scenario is similar to the previous one in that it describes someone making a remix request. However in this case the requester does not have aspirations to be a remixer but is simply looking for a particular kind of media to consume. Her request differs from the previous one in that she specifies only a song to be used, not a detailed creative vision. Posts like this, usually specifying a musician or song name and maybe the name of an anime series, are also extremely common on the AnimeMusicVideo.Org forums.

Kira logs on to AMV.Org and searches once again to see if there are any new videos using Black Remembrance songs. Nope, no such luck. Why don't these people understand how awesome Black Remembrance is? She posts for what seems like the hundredth time to the "AMV Suggestions" forum:

We need MORE Black Remembrance videos! Especially "Rain Flowers." If anyone of you guys uses this song in an AMV, you will be my all-time HERO! Do it!

A few minutes later she gets a reply:

Black Remembrance is lame. If you like them so much, why don't you make your own video?

She really is sick of hearing that. She doesn't want to make videos, she wants to watch them. Why can't someone just make a Black Remembrance video? It would be so cool!

Collaboration model

As described in the scenarios above, participants in the AMV community collaborate to produce and enjoy remixed works. This collaboration includes not just the personas above but also the *anime* and music industries, as well as an array of less involved participants such as the software companies that develop and sell editing software and the Internet hosting companies that host AMV websites. Jenkins has noted that fan production of this sort constitutes an example of what sociologist Harold Becker calls an "art world."¹⁶ An "art world," according to Becker, is "an established network of cooperative links" which produces artistic works.¹⁷ Figure 4 shows a subset of these cooperative links as they currently exist in the AMV community.

As the diagram shows, Daniel and Takuya (and, perhaps less purposefully, the *anime* and music industries) cooperate to provide music videos for Kira's enjoyment. But although Kira plays an important role in appreciating and supporting Daniel and Takuya's work, her attempts to collaborate more actively by providing input in the form of requests are stymied. The next section will consider how these patterns of cooperation might be changed by introducing new mediating technologies.

¹⁶ Henry Jenkins, *Textual Poachers* (New York: Routledge, 1992), 46-47.

¹⁷ Harold Becker, *Art Worlds* (Berkeley: University of California Press, 1982), 34-35.

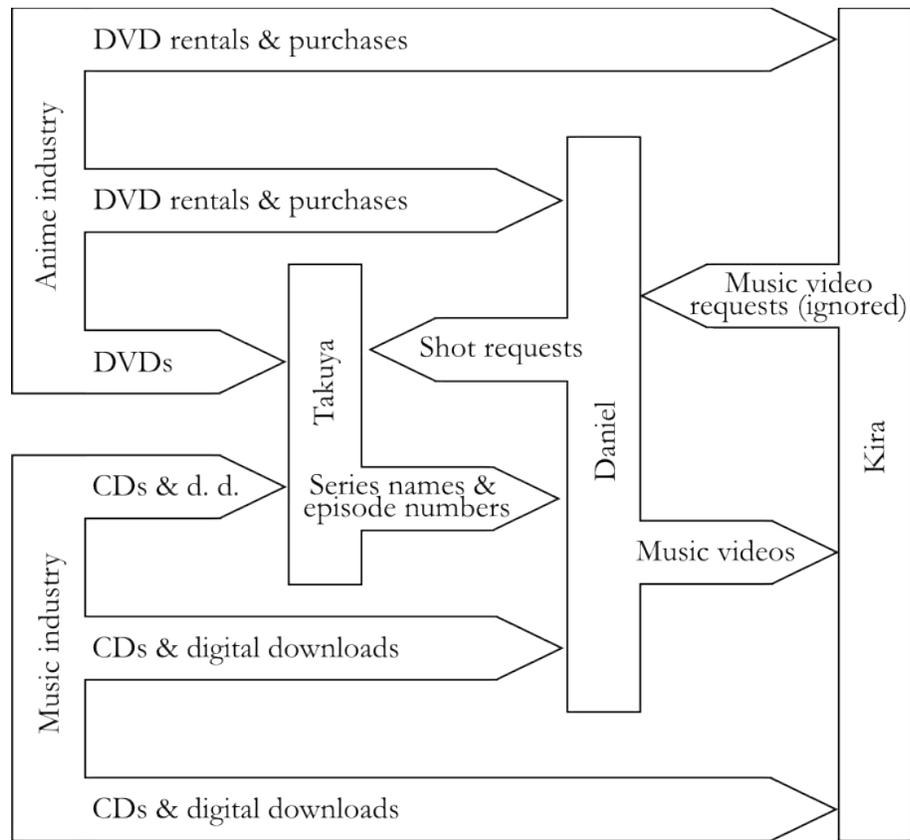


Figure 4. Cooperative links in the AMV community.

Proposed mediating technologies

In the previous section we looked at the different kinds of participants in the AMV community and each participant's "bundle of tasks." Becker claims that participants in an art world will seek to "get rid of tasks they regard as tiresome" and "add tasks that are more interesting, rewarding, and prestigious."¹⁸ Successful mediating technologies must therefore help people in these efforts.

Three mediating technologies are proposed. The first is a media-on-demand network for delivering original music and anime from the content industries to Daniel, Takuya, and Kira. Next is a shot metadata repository for mediating interactions between Daniel and Takuya. The final component is an alternative system for generating remixes customized to Kira's specifications, based on rules provided by Daniel and Takuya.

¹⁸ Ibid., 11-13.

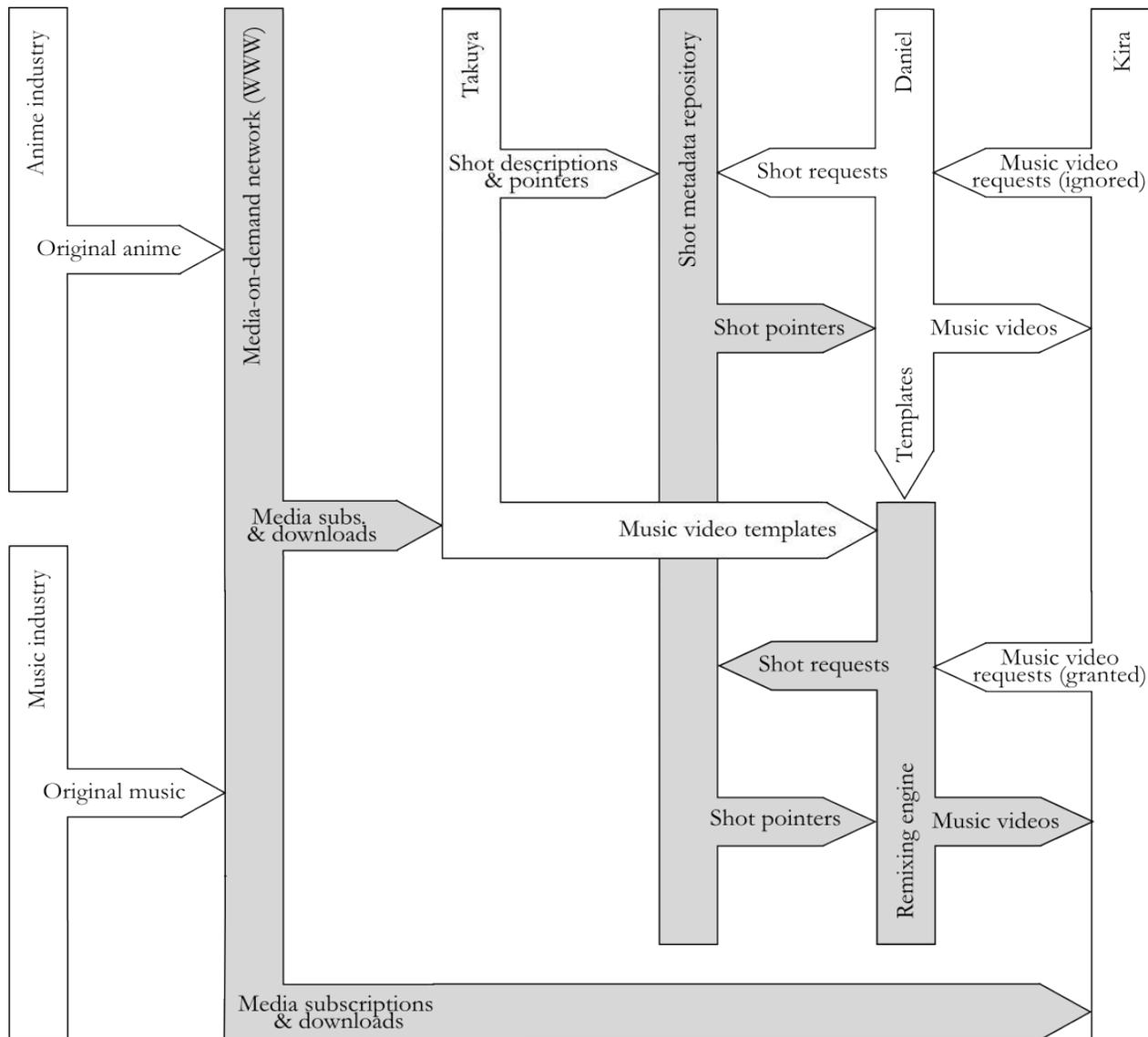


Figure 5. Proposed mediating technologies. (Media flows from WWW to Daniel omitted for clarity.)

Media-on-demand network

Several of Daniel's most tiresome tasks could be eliminated by replacing the current system of content distribution via physical CDs and DVDs with a fully digital, on-demand content distribution network. Technically, such a system is already in place: there are a number of options for downloading music on demand, and fans can obtain virtually any desired *anime* titles through peer-to-peer distribution networks.¹⁹

But the existing system is not ideal: peer-to-peer distribution of anime usually violates copyright law, and fans frown on the illegal distribution of content that can be obtained legally through other means.

¹⁹ See for example AnimeSuki.com (<http://www.animesuki.com/>), DownloadAnime.org (<http://www.downloadanime.org/>), and BoxTorrents (<http://www.boxtorrents.com/>).

Furthermore, the existing systems use on a multiplicity of incompatible protocols and formats, making the use of downloaded content more of a hassle than relying on old-fashioned physical media. Finally, video content distributed on peer-to-peer networks is usually compressed in such a way as to make frame-accurate editing very difficult or impossible, since not every frame is a keyframe.

An ideal media network would provide video files in a unified format comparable to DVD. Such a network would likely be enthusiastically embraced by Daniel, as it would eliminate the need for ripping DVDs to his hard drive (considered “somewhat tedious” or “very tedious” by 56% of respondents in the survey). More importantly, however, such a system could enable a radical change in the way remixes are distributed.

If the media network provided not only full audio or video files, but media fragments (specified by temporal indexes into media files), and these fragments were fine-grained enough (down to the level of a single frame of video), Daniel could distribute his music videos as a list of fragment pointers, possibly with some additional instructions for applying effects and adjustments at play time. Kira could load this “super playlist” into her computer/entertainment system, which would then assemble the music video and play it (assuming Kira had access rights to the necessary source media).

This approach to the distribution of remixes could have enormous consequences for the creation of AMVs. Moving from the distribution of remixed media files to the distribution of edit decision lists might lift the practice out of the ghetto of quasi-legality where it currently resides. AMV editors are currently creating and distributing derivatives of copyrighted works, which is illegal. The *anime* industry tends to look the other way, since it is not in their interest to start suing their best customers, but this does not change the fact that the AMV editors are breaking the law and must live in fear that they will be forced to destroy their work.

The pointer-based approach described above may seem like a technical end-run around copyright law. Would this practice not be considered distribution of derivative works? There is some indication that it would not be. The Family Entertainment and Copyright Act of 2005 modified copyright law to allow “the making imperceptible... of limited portions of audio or video content of a motion picture, ...or the creation or provision of a computer program or other technology that enables such making imperceptible” as long as “no fixed copy of the altered version of the motion picture is created.”²⁰ One can imagine media player which, upon receiving a music video edit decision list, loads the source content and “makes imperceptible” all content except that which the remixer has selected.

On the other hand, the law does state that only “limited” portions may be excised, and furthermore prohibits “the addition of audio or video content that is performed or displayed over or in place of existing content.” These restrictions may mean that a pointer-based approach is doomed as well, in which case the only hope for legal remixing is to persuade content owners to relinquish some of their rights by licensing their content under a license like the Creative Commons Sampling License.²¹

A media distribution network like the one described may arrive sooner rather than later. Gotuit Media is already selling a video-on-demand system for digital cable networks which enables users “to bookmark points in a program and to share saved segments with other subscribers.”²² Ideally, however, such a system would not reside on a proprietary network like Gotuit's, but on the Web, where it could be fully integrated

20 Family Entertainment and Copyright Act of 2005, Public Law 109-9, 109th Cong., http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_cong_public_laws&docid=f:publ009.109.

21 Creative Commons, “About the Sampling Licenses,” <http://creativecommons.org/license/sampling>.

22 Gotuit Media, “Products and Services,” http://www.gotuit.com/products_services/.

with the existing wealth of information available there. The Annodex project is currently developing standards and tools for just such integration, enabling hyperlinking and on-the-fly composition of audio and video served over the Web.²³ Other research efforts are more ambitious, seeking find ways to enable the full range of media fragment integration necessary to provide through the Web the functionality expected by AMV editors, including spatial compositing.²⁴

Standardized ways of pointing to and retrieving fragments of media are important, but we still need ways to easily find fragments of interest and specify ways to recombine them. In other words, we need metadata to associate with our media fragment pointers.

Shot metadata repository

The interactions between editors like Daniel and experts like Takuya could be enhanced through the construction of a shared database for associating *anime* shot content with higher-level descriptions of that content. This database would not replace the (relatively) real-time requests and responses that currently occur between editors and experts, but would augment them by allowing enthusiasts to capture their contributions in a way that makes them more useful to others in the future.

Enthusiasts could contribute annotations just for fun, or in the process of doing something more creative, like creating a storyboard for a suggested AMV. Remixers could add annotations as part of their creation process—any labels created for their own use while organizing their shots would automatically be contributed to the shared metadata repository. Even if remixers do not label their shots, metadata generated through their activity, such as how often particular shots were used in remixes and which other shots they were often juxtaposed with, could be added to the repository.

Such a database would allow remixers to search for shots of interest and obtain pointers to those shots without having to painstakingly trim each shot out of the surrounding video (a task 47% of survey respondents considered “somewhat tedious” or “very tedious”). This kind of functionality was of great interest to the interviewees:

I'd love the ability to type in, "Scene with Shinji looking surprised," and have it give me a list of all of the scenes.

We both would love the ability to go in and chop up the full episode or full hour and half of video into individual scenes, at scene changes or scene cuts. And then to be able to take those individual cuts, and be able to search them on a conceptual basis, in an automated fashion, would be a god-send.

[T]hat person needs to find characters eating pancakes, and that person can go to your database, and then that database will tell them, "Oh if you go to Trigun, episode..." Ahh... that's a good idea. That's a very good idea.

The survey respondents likewise responded positively, with 54% of respondents calling “an anime search engine which would allow searching for specific shots or scenes by character, location, mood, or other criteria” a “good” or “great” idea.

An important aspect of the metadata repository would be the linking of metadata to the identity (or identities) of the user who created it. A reputation system which allowed contributors to turn their cultural capital (expert knowledge of *anime* content) into social capital (widespread recognition and validation of

23 Silvia Pfeiffer, Conrad Parker, and Claudia Schremmer, “Annodex: A Simple Architecture to Enable Hyperlinking, Search & Retrieval of Time-Continuous Data on the Web,” in *Proceedings of 5th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, 87-93 (Berkeley, California: ACM, 2003).

24 Lloyd Rutledge and Patrick Schmitz, “Improving Media Fragment Integration in Emerging Web Formats,” in *Proceedings of the International Conference on Multimedia Modeling 2001 (MMM01)*, 147-66 (Amsterdam: CWI, 2001).

their knowledge) could go a long way toward increasing the prestige of knowledgeable enthusiasts. The importance of such accreditation among remixers is clear in the AMV community, where “stealing someone's idea” without crediting them is a serious norm violation.

Finally, the metadata repository should store not just high-level, human-provided metadata but also whatever low-level metadata can be automatically provided by machines. One interviewee said that he would like to search for scenes by dominant color, an attribute more easily provided via color histograms than manual annotation. Other kinds of attributes like motion vectors may also be more suitable for automatic annotation.

A rich repository of annotations and pointers to media fragments and a system for querying that repository could make possible a number of applications, including a system for automatically generating AMVs.

Remixing engine

While the AMV community was described above as a cooperative network, it is still rather hierarchical. A significant prestige gap exists between remixers and other members of the community. Successfully creating an AMV involves mastering a number of difficult concepts and technologies, and though in theory anyone can find the help they need through the resources at AMV.Org, in practice there seem to be a large number of people interested in making AMVs but unwilling or unable to make the necessary commitments to do so.

The interactions between these “wannabes” and the rest of the community could potentially be improved by creating a music video template database, containing templates or sets of rules for combining annotated media content, and an engine for executing these rules to produce customized media. Such a database would give enthusiasts like Takuya an outlet for their creative aspirations, and provide newbies like Kira with the power to exert some control over the creation of the media they see and hear. In other words, such a system would allow both semiotic and enunciative producers to take steps towards becoming textual producers.

There are a number of examples of such systems in the multimedia research community. Perhaps the earliest such system was Bloch's tool for generating film sequences using a set of rules for how shots can and cannot be combined.²⁵ More recently, Bocconi and Nack's VOX POPULI system shows that rules for argumentation can be used to generate video sequences which fulfill specified rhetorical aims.²⁶ Systems for automatically generating music videos have been developed as well. For example, a system developed at FX PAL uses low-level features obtained through machine perception techniques to select “good” shots from video and synchronize them to an audio soundtrack.²⁷

But while the automatic generation of AMVs seems possible from a purely technical point of view, there may be some issues with integrating such technologies with a social system like the AMV community. One interviewee expressed a desire for a semi-automatic editing tool:

The perfect AMV editing machine... would take the music that I wanted, it would extract the exact scenes I wanted, and put them to the exact lyrics and words.

25 Gilles R. Bloch, “From Concepts to Film Sequences,” in *Proceedings of RLAO*, 760-67 (Cambridge: RIAO, 1988).

26 Stefano Bocconi and Frank Nack, “VOX POPULI: Automatic Generation of Biased Video Sequences,” CWI technical report INS-E0405 (2004).

27 Jonathan Foote, Matt Cooper, and Andreas Girgensohn, “Creating Music Videos using Automatic Media Analysis,” in *Proceedings of ACM Multimedia* (2002).

But 63% of survey respondents considered “an editor which could take a specific song and specific anime content and automatically create a rough edit of an anime music video” to be a “bad” or “terrible” idea, with the majority rating it “terrible.”

It is easy to see why remixer would dislike the idea of being replaced by a machine, but the number of unanswered requests for music videos observed on discussion forums suggests that there is far more demand for remixed material than can currently be met. An automated system could possibly satisfy this demand. However, such a system must be viewed as providing a lower-quality alternative to a handmade music video and not as a replacement for a human remixer. If remixers consider the system to be more of a “toy” to allow unskilled fans to play with media, they may embrace it, as it could free them from the constraints of popular appeal²⁸ and allow them to focus on more artistic visions.

“To-be” scenarios

The scenarios in this section show how the systems described above might change the patterns of interaction in the AMV community.

Scenario 1B: The remixer

This scenario suggests how the time requirement for creating a video could be cut down from several weeks to as short as a few hours, by eliminating or automating tedious setup work and integrating powerful search functions into an editing tool.

Daniel is in his room working on his computer and listening to his favorite streaming music station. A song comes on that catches his ear. He clicks the “Show lyrics” button on his media player. This causes the MSMDX plugin which he has installed in his media player to send the song's ID to the MSMDX lyric service, which responds with the lyrics for the song and precise timestamps indicating when in the song each word is sung. A small window opens and the lyrics to the currently playing song begin scrolling by.

*I wonder why we waste our lives here
When we could run away to paradise
But I am held in some invisible vise*

Perhaps it's senior-year burnout, but the lyrics strike a chord with Daniel. He wonders if maybe this song would be a good candidate for a music video and decides to investigate further. He has the MSMDX plugin do a lookup for related *anime*.

The plugin again sends the song ID to the MSMDX lyrics service, but this time it instructs the service to return statistically improbable phrases rather than the full set of lyrics. These are phrases that appear in the lyrics of this song but which are not commonly found in the lyrics of other songs. In this case the MSMDX lyrics service identifies the phrases “invisible vise” and “bars come down around me” as most significant and returns them to the plugin. The plugin then uses these phrases to send a query to the MSMDX shot database. The query returns a list of short *anime* clips, which appear on Daniel's screen as a list of keyframes. Most of them depict *anime* characters in various states of bondage or confinement—several in jail cells, a few in handcuffs, one which appears to be strapped into some kind of medieval torture machine.

Looking over the keyframes, Daniel decides that these clips, while interesting, don't really fit the mood

²⁸ As one interviewee put it: “*Anime* fans are fickle, they're always there to look for a laugh rather than look for art or look for something that's actually just kind of a cool music video... there are things that are very good but not considered successful because not many people watch it.”

of the song—a little bit too much of an S&M flavor to them. When he listens to the lyrics, he thinks about being trapped in school—what about *anime* showing kids having a bad time at school? He opens the MSMDX shot database search interface and types in “bad time at school”. Again he sees a list of keyframes, this time depicting a variety of school-related situations, none of them pleasant-looking.

He skips over the first couple of results—the first result is from *Urusei Yatsura*, which he considers overused, and the second is from *Iron Intellectrix*, which he doesn't really like—but the third one, a long shot of a kid alone in a classroom, looking scared, intrigues him. The description says it's from *Cromartie High School*, which he's seen a few episodes of and liked. He clicks the keyframe for a closer look. A ten-second clip plays, showing the aforementioned young man, whose name according to the expanded description is Takashi, alone in the classroom as the camera pulls back into the hallway and the door slams.

Daniel likes what he sees. He opens up his video editor and drags the clip over to it. He also drags over the song, which causes a waveform to appear on the editor's timeline. Thanks to the MSMDX editor plugin he has installed, he also sees the lyrics running underneath this waveform, each word placed at the precise point in time when it is sung. Daniel positions the clip of Takashi in the classroom above the phrase “Is this the room I'll live my life forever.”

Now Daniel runs another shot search, this time using the search capabilities of his MSMDX editor plugin (rather than the MSMDX media player plugin he was using earlier). He requests every shot from *Cromartie High School* that shows Takashi either at school, or in a generic setting. The request returns a long list of clips, which he begins examining one at a time, playing the clips that interest him, and dragging the relevant ones to various places on the editor timeline.

Soon Daniel has a pretty decent music video skeleton in place, but he's finished looking at all the *Cromartie High* clips and there are still some gaps on the timeline. He decides to return to his earlier “bad time at school” query again and peruse those results a bit further. A keyframe showing a group of menacing bullies gathered in a schoolyard catches his eye. The description says it's from *Great Teacher Onizuka*—he's not familiar with that one. He clicks the link for more information on the source and discovers that *Onizuka* is a series about a young slacker who becomes a teacher at a tough high school. That sounds relevant. He watches the clip and decides it will fit in well with the *Cromartie High* footage, so he drags it onto the timeline. He runs another query, this time restricting it to shots from *Onizuka* in a school setting. He is glad to see that he gets a healthy list of results—surely he can find enough material to finish his video here.

Scenario 2B: The enthusiast

This scenario suggests how a music video storyboarding tool could allow enthusiasts to annotate shots and create templates for music videos, which could then be handed over to an remixer for completion or used by an customized media engine (see Scenario 3B) to generate music videos.

It's Friday evening and once again Takuya is playing with the MSMDX storyboarding tool. He's just about to finish another take on his planned *Iron Intellectrix* video, a tearjerker telling the tale of Tran and Shiv's ill-fated attempts to rescue their father.

The tool's interface shows keyframes from the shots he has selected so far, alongside lyrics from the song he has selected (“Goodbye” by Disappearance) and his notes on each shot's content and why it was selected. He notices one shot has hasn't labeled yet, so he selects it and types “Tran/Shiv digging up the grave.” Each time he adds a note like this his description is uploaded to the MSMDX shot database, along with the ID of the original *anime* episode from which the shot was taken, and its beginning and ending

frame numbers.

Next Takuya decides that the order of shots isn't really so great, so he has the tool reorder them by the order in which they occur in the original show. He decides to test out his creation and see how it flows. The song begins playing, and as each lyric begins its associated shot plays in sequence. It's not exactly a fully formed music video—more of a template—but Takuya is satisfied. Maybe one of the editors will pick it up and smooth out the syncing and transitions. He uploads it to the MSMDX template database.

Scenario 3B: The newbie

This scenario describes a system for automatically generating music videos.

Kira logs on to AMV.Org and searches once again to see if there are any new videos using Black Remembrance songs. The search returns no results, and the system offers to generate a music video for her. She accepts, and is presented with a list of Black Remembrance songs. She selects “Rain Flowers.” She decides not to choose a specific *anime* series and clicks “Show Video.”

The system analyzes the length and tempo of “Rain Flowers” and searches through its music video template database to find an appropriate set of rules for setting shots to it. Once it finds a template it searches through the shot database to find shots which fit the template. It knows to avoid shots from Dragonball Z, as Kira has in the past stopped watching videos featuring Dragonball Z content an after an average elapsed time of ten seconds. Finally it finds enough shots to fill the template, so it applies some transitions, synchronizes the action in the shots to the beats of the song, and the video is complete.

Kira's media player opens and the video begins playing. *Awesome*, she thinks. She's not familiar with the *anime*, but it looks pretty cool, and it suits the song well. At one particularly cool-looking shot she clicks the media player display, which pauses the music video and opens a new window showing that shot in the context of the original *anime* series from which it was taken. She watches for a bit, decides she likes it, and indicates that she would like to download the series to watch later. After it begins downloading she returns to the original display and resumes watching her new Black Remembrance music video.

Matching shot requests to shot descriptions

A key component in the scenarios above is a shared metadata repository which holds pointers to specific spatio-temporal fragments of media along with descriptive metadata for the content of those fragments. Such a repository can potentially ease collaboration between Daniel and Takuya by allowing their collaboration to happen incrementally and asynchronously, rather than requiring approximately real-time interaction.

But real-time interaction has its advantages, as it allows Daniel to communicate the context of his request to Takuya, and both Daniel and Takuya can ask questions to resolve problems such as ambiguity in the request. If Daniel and Takuya are collaborating through a shared metadata repository, these problems are not as easily addressed, and some new problems are introduced. In particular, since Takuya is now providing descriptions of content not in response to a request from Daniel, but before Daniel has even made a request, there are new coordination problems.

First there is the well-known problem of coordinating the vocabulary of the metadata provider (Takuya) with the vocabulary of the information seeker (Daniel). For example, Takuya may describe a character in a shot as a “bad guy” while Daniel may request shots portraying “villains.” A controlled vocabulary with lists of synonyms for each term in the vocabulary is the traditional solution to this problem.

More worrisome is the possibility that Takuya will choose to describe attributes of content unrelated to the kinds of attributes Daniel specifies in his request. For example, Takuya may choose to spend all his time describing the lighting in a shot, neglecting to provide any information about the action it depicts. If Daniel is interested in action, Takuya's annotations will be useless to him. This could be addressed by requiring Takuya to annotate all possible facets of a given shot, but this solution makes the cost of contributing unacceptably high (to use Benkler's term, contributions are no longer granular enough). An experiment was conducted in order to investigate the seriousness of this potential problem.

Request analysis

In order to better understand how remixers requested content, 220 requests for anime shots made on the AMV.Org "AMV Suggestions" forum from July 2002 to April 2005 were analyzed using a two-pass approach.²⁹ First 100 requests were examined and categorized according to the attributes used in specifying the request. The set of attributes was not decided ahead of time, but was developed in parallel with this examination, and thus was based on the actual data. Next this coding scheme was used to categorize the full set of 220 requests. The results are shown in Table 1.

²⁹ Morten Hertzum, "Requests for Information from a Film Archive: A Case Study of Multimedia Retrieval," *Journal of Documentation* 59, no. 2 (2003): 173-74.

	Attribute	Examples	Requests containing attribute	
			n	%
Appearance-related	action	breakdancing; gambling; trying to commit suicide	89	40
	object	cool shoes; straitjacket; pancakes	64	29
	body/face	adult female; werewolves; guys that look like girls	59	27
	setting/era	a bar; in the darkness; a jungle	29	13
	number	two characters; large group	8	4
Subject-related	plot/theme	one character dumping another; heroic characters working toward a common good	26	12
	mood/emotion	silly; psychedelic; sad	19	9
	role	villain; father; maniac	17	8
	genre	sci-fi; romance; <i>yaoi</i>	11	5
Specific names	cast member name	Vash; Tenchi	20	9
	film/series name	<i>Escaflowne</i> ; <i>Trigun</i>	20	9
Production-related	camera/animation style/etc.	camera scrolls up; zoom in or out; loopable	9	4

Table 1. Shot request attributes and their frequencies.

Requests specified two different kinds of attributes on average, with a combination of action and object being the most common (14% of requests). The most commonly specified attributes overall were appearance-related, especially actions depicted, specific objects, and specific physical attributes of bodies or faces. Specific names of characters, films, or TV series were less common than expected.

One surprise was the number of requests that specified plots or themes, since plots and themes are things a remixer can change through the editing choices he or she makes. In some cases these requests may have been due to the remixer not realizing this and assuming that, in order to create a music video about a boy and a girl breaking up, source material from an anime with the same plot is needed. On the other hand, this may reflect the fact that remixers often seek to comment on original storylines known to their audiences. For example, a remixer could be creating a tribute to classic break-up stories, in which case footage is needed to refer to plot or theme of the original narrative, not to create a new narrative.

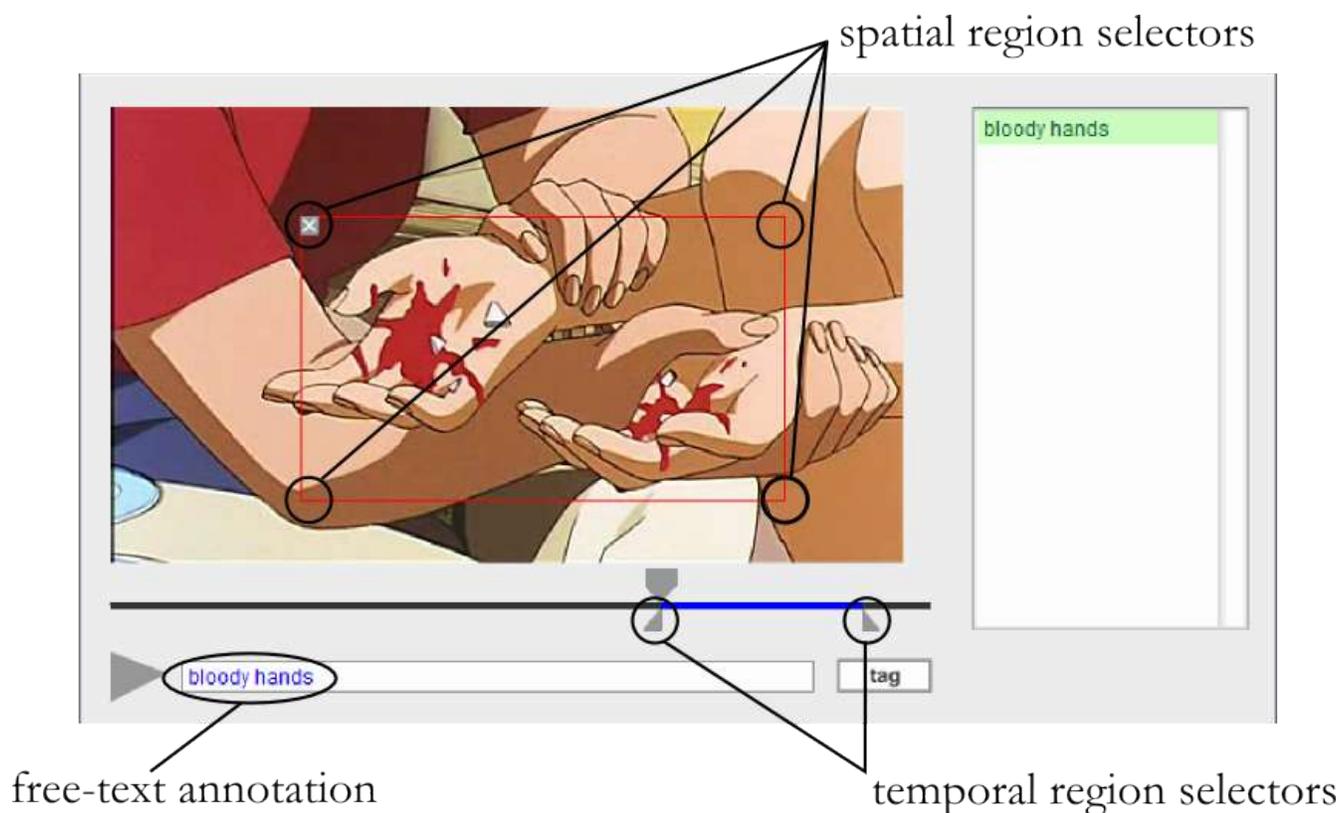


Figure 6. Web-based video annotation tool.
Annotation experiment

Next an experiment was devised to provide some data on how enthusiasts might choose to describe content, given a tool for doing so. A simple media player application was created which enabled users to select temporal and spatial regions of two one-minute anime clips and add free-text annotations (see Figure 6). The application was implemented using Macromedia Flash so that it could be accessed through a standard web browser. Annotations and their corresponding selected regions were stored in a relational database as they were created. Users could see a list of all annotations created and could select annotations in the list to see the associated regions highlighted in the media player. Annotations could not be removed once they had been added.

The first one-minute clip was a scene from the English-dubbed USA release of *Cowboy Bebop: Knockin' on Heaven's Door* (2001),³⁰ a popular anime film used as source content for 568 music videos listed on AMV.Org. This scene was selected because of its high action content, relative lack of dialog, and the likelihood that it would be well-known to fans. The second clip was a scene from the English-dubbed USA release of *Perfect Blue* (1997),³¹ a somewhat less well-known film used as source content for 250 music videos listed on AMV.Org. This scene featured mainly a dialog between two characters.

60 survey respondents who had provided contact information and expressed interest in participating further in the project were invited to use the tool to annotate the two clips. Invitees were given no guidelines on what or how to annotate beyond basic instructions on how to use the annotation tool. The

30 Original Japanese title is *Cowboy Bebop: 天国の扉*. IMDb, "Cowboy Bebop: Tengoku no tobira," <http://www.imdb.com/title/tt0275277/>.

31 IMDb, "Perfect Blue," <http://www.imdb.com/title/tt0156887/>.

instructions simply asked users to “Add as many tags as you think are necessary to describe the scene.”

25 respondents accepted the invitation and created a total of 244 annotations for the two clips (see Table 2). 210 (86%) of these annotations specified particular temporal regions of the clip. In most of the cases where a specific region was not specified, it seems that this was due to user misunderstanding, since the annotation text seemed to refer to specific shots rather than the scene as a whole (although there were a few annotations which described the scene as a whole). Only 16 (7%) of the annotations specified spatial regions. This low number was probably due to the fact that the instructions did not make it sufficiently clear that it was possible to make spatial selections. Since, unlike temporal selecting, this is not a common feature of tools like video editors with which the annotators were familiar, it is likely this functionality was overlooked.

	Cowboy Bebop	Perfect Blue
Percentage of annotators who reported having seen the source film	64%	48%
Total number of annotations	161	83
Percentage of annotations specifying temporal regions	84%	90%
Percentage of annotations specifying spatial regions	6%	8%
Median number of annotations per annotator	5	3

Table 2. Annotation statistics for the two anime clips.

A wide variety of annotation styles were observed. Some annotators wrote long descriptions consisting of complete sentences:

Nina's so stressed out and confused that she breaks her cup in her hands and isn't phased by it but meerly shows her state of mind by wondering if it was real. It shows that she's struggling with reality in a very dramatic and scary fashion...like a psychopath ^_^ (Then again, she IS psychotic...and the stalker...THE STALKER! AHH!!)

Others used short, “tag” style annotations. One annotator in particular eschewed the use of spaces, creating tags like “StrutLowAngle.”

The descriptions were coded using the set of attributes developed from the shot request analysis (see Table 3). Descriptions specified two different kinds of attributes on average, with a combination of action and object being the most common (14% of descriptions). The most commonly specified attributes overall were appearance-related, especially actions depicted, specific objects, and specific physical attributes of bodies or faces. Specific names of characters were mentioned quite often, though the names of the films from which the clips were taken were not.

Three new attributes were observed that had not been encountered in the shot requests. Three annotators described a shot focusing on a woman's bloody hands as “stigmata” or “Christ symbolism.” This was considered sufficiently different from “plot/theme” to warrant a new attribute, “symbolism.” A number of annotators also described the audio tracks of the clips, something not encountered in the shot requests (presumably since AMV editors do not generally use the audio tracks from *anime* content). Finally, one annotator described how certain clips might be used for creating AMVs, for example annotating spatial regions which (he claimed) could be easily altered using Adobe Photoshop.

	Attribute	Examples	Annotations containing attribute	
			n	%
Appearance-related	action	escape jump; wakes up	109	45
	object	teacup; truck	58	24
	body/face	looks confused; bloody hands	42	17
	setting/era	highway; bedroom	28	11
	number	<i>never used</i>	0	0

	Attribute	Examples	Annotations containing attribute	
			n	%
Subject-related	mood/emotion	creepy; strange	52	21
	plot/theme	Faye realizes she is tracking the wrong person; Mima begins questioning reality	29	12
	role	evil guy; mom	4	2
	symbolism	stigmata	3	1
	genre	action thriller	1	0
Specific names	cast member name	Vincent; Rumi	79	32
	film/series name	<i>Cowboy Bebop</i> ; <i>Perfect Blue</i>	4	2
Production-related	camera/animation style/etc.	nice drawing; close-up	40	16
	audio/dialog	some quiet creepy music; “Is it real?” [quoting dialog]	15	6
	remix suitability	ripe for Photoshopping; made for AMV	4	2

Table 3. Shot description attributes and their frequencies. New attributes are shown in bold.

Comparison

The frequency distributions of attributes for the shot requests and shot descriptions were quite similar (see Figure 7). As mentioned above, both requests and descriptions were most likely to specify actions depicted, with a combination of action and object being the most common type of annotation.

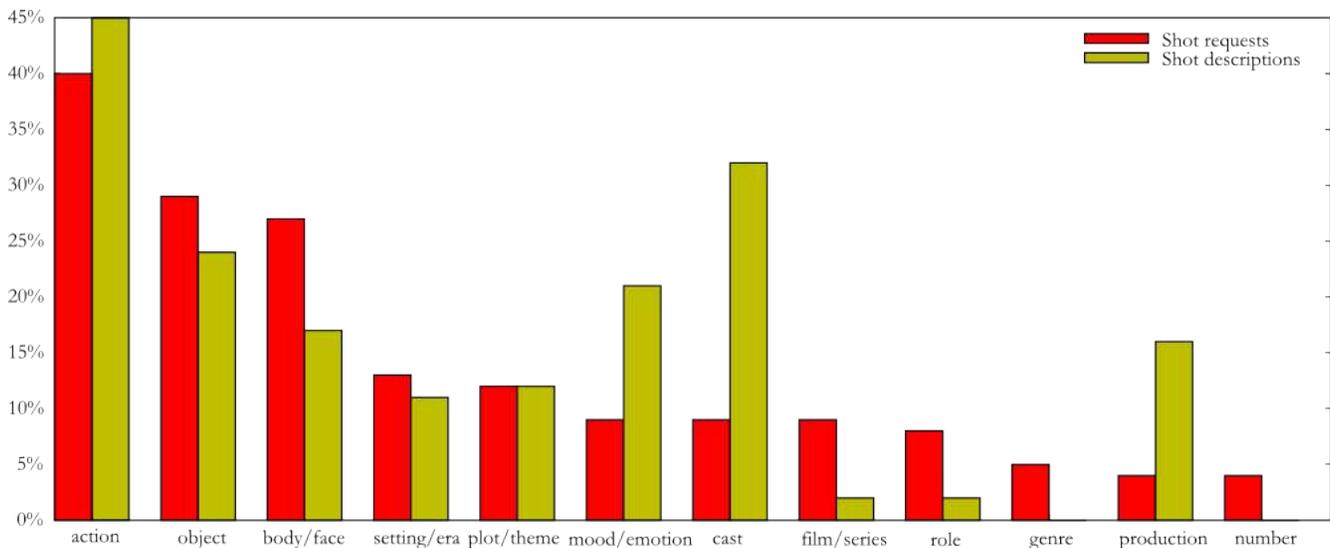


Figure 7. Shot request and shot description attributes and their frequencies.

Some potential sources of coordination problems can be seen in the differences between the percentage of requests specifying physical attributes of bodies or faces (27%) and the number of descriptions specifying these attributes (17%), and between the percentage of requests specifying roles (8%) and the number of descriptions specifying these attributes (2%). (There is a similar gap for specific mentions of film or TV series names, but this is not a problem as this metadata can be assigned fairly easily through means other than manual annotation.)

These differences point to the need for a domain ontology which links names of characters (“Vincent”) to those characters' physical appearances (“tall,” “male,” “dark hair,” “tattooed”) and the roles they play (“villain”). Such an ontology would enable the expansion of descriptions specifying names to also include these associated attributes. Since annotators seem to specify character names quite often, this approach could alleviate the coordination problems.

Conclusion

This study provided detailed data on how a particular community describes, looks for, and remixes media content, which was used to develop a framework for understanding how different members of this community interact with each other and with the media they love. Based on conversations with members of this community, observations of their online behavior, and the opinions they expressed when surveyed, it seem likely that this community will use collaborative, metadata-driven tools for finding, sharing, and remixing audio and video, if these tools can be designed to fit into their existing practices.

In the next phase of this project, the data gathered will provide “real world” use cases and specific examples for testing the suitability of representation and query languages. The result will be a far more robust design that is much likelier to achieve acceptance. In the next section the findings presented here are distilled into a set of requirements for this design.

Initial MSMDX platform requirements

The requirements enumerated in this section are based partly on the findings presented in the previous section, and partly on “best practices” observed by other successful community metadata projects.

A successful system for catalyzing community production of multimedia metadata will have several general qualities:³²

- **Simplicity.** MSMDX interfaces must present low cognitive costs to users in order to encourage fine-grained, low-effort contributions.
- **Community.** MSMDX must give users opportunities to build cultural capital (by increasing their knowledge about media of interest) and social capital (by sharing this knowledge with others).
- **Expressiveness.** MSMDX must have a data model rich and flexible enough to allow users to make whatever statements they wish about media, including statements we can't currently imagine.
- **Openness.** MSMDX must allow users to create metadata for any media without having to obtain permission to do so, and no registration should be required beyond what is necessary to associate a stable identity with annotations.
- **Integration.** MSMDX should integrate into existing social practices and be usable with existing tools.
- **Platform independence.** MSMDX must allow users should be able to create annotations for a piece of media regardless of differences in platforms or proprietary media formats.

The last few requirements entail that MSMDX must be developed using open standards, ideally with an open source reference implementation. However, since MSMDX must also interface with a variety of proprietary systems, it should be licensed in such a way as to make utilization of its code as easy as possible for both open and closed software developers.

User requirements

MSMDX is intended to serve a number of different users, who can be organized according to the roles they play as producers and consumers of media and metadata:

- **Original creators.** These users include the traditional content industries (the record industry, film studios, etc.) as well as independent creators (indie musicians, Flash animators, videobloggers).
- **Enthusiasts.** These are hard-core fans, who passionately and methodically pursue greater knowledge of the media they love.
- **Newbies/Casual fans.** These are ordinary fans looking for casual entertainment that requires no great expertise.
- **Remixers.** These are creative amateurs who use media as raw materials to produce new works.

These roles are only rough categories; many users will fill more than one role, or play different roles at different times. Figure 8 shows how these different users cooperate to exchange media and metadata using the MSMDX platform.

³² See also Robert Wilensky's criteria for an online annotation system that can support spontaneous collaboration in “Digital Library Resources as a Basis for Collaborative Work,” *Journal of the American Society for Information Science* 51, no. 3 (2000): 229.

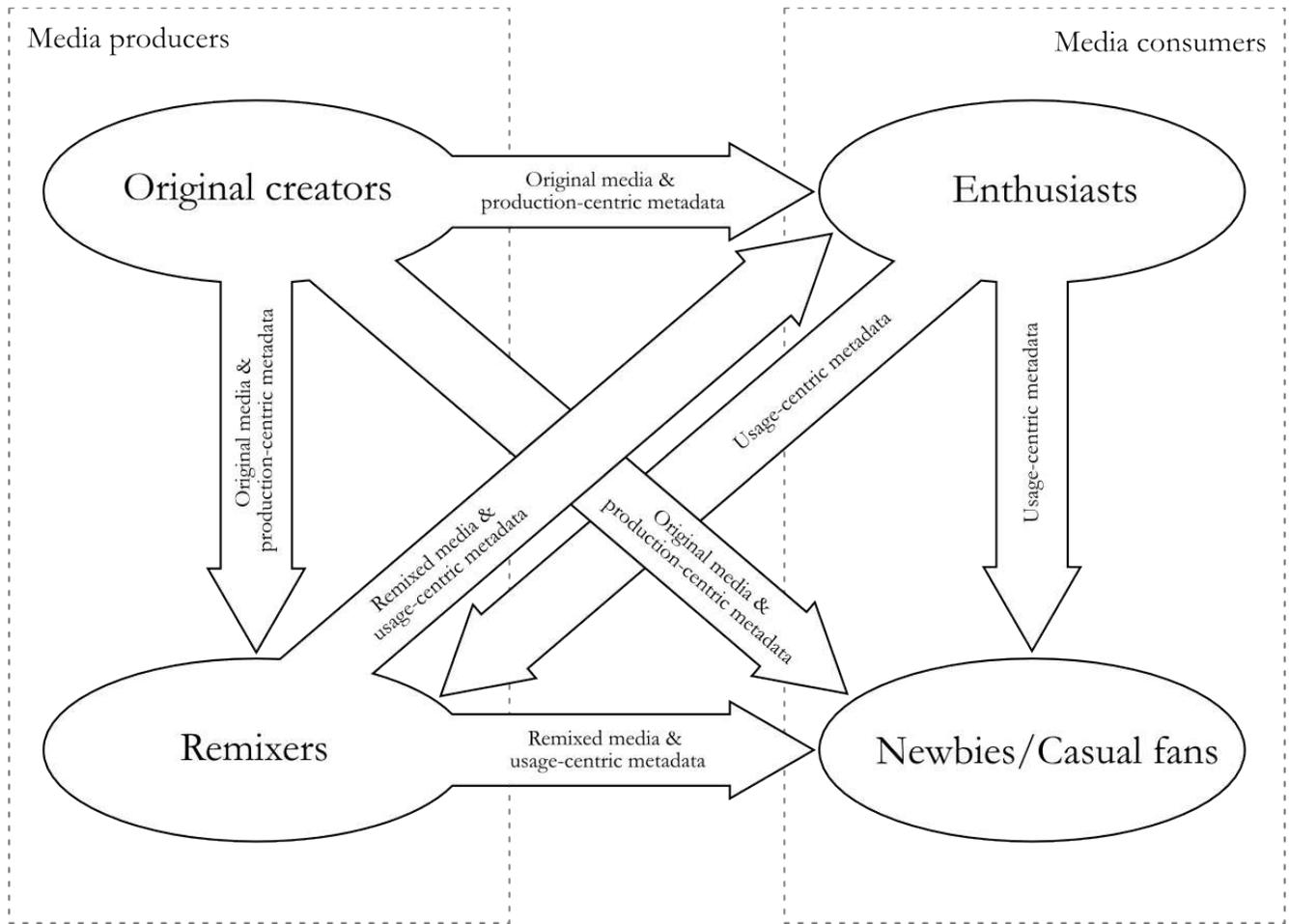


Figure 8. Flows of media and metadata between MSMDX users.

Use cases

Use Case	Examples	Target Users	Priority
Describe audio file	Rate a song; categorize a song; comment on a song	Newbies	1
Describe video file	Rate a TV episode; comment on a TV episode; specify the cast or crew of a film	Newbies	1
Describe image file	Add a caption to a photograph; comment on a drawing	Newbies	1
Describe audio fragment	Comment on a drum break; specify who is playing a solo	Enthusiasts	2

Use Case	Examples	Target Users	Priority
Describe video fragment	Describe the setting of a shot; label a single moving object in a shot	Enthusiasts	1
Describe image fragment	Select and name a single face in a group photograph	Enthusiasts	1
Find audio file	Search for a song by mood; search for a radio interview by interviewee name	Newbies	1
Find video file	Search for a TV episode by actor; search for a feature-length film by name	Newbies	1
Find image file	Search for a photograph by event; search for a comic strip by topic	Newbies	1
Find audio fragment	Search for a drum break; search for a “sound bite”	Remixers	2
Find video fragment	Search for a video shot by depicted characters; search for an object moving in a particular way; search for a background shot	Remixers	1
Find image fragment	Search for a particular person or object within a photograph	Remixers	2
Find timed text	Get lyrics for a song; get transcript of a TV episode	Enthusiasts	1
Create timed text	Transcribe lyrics of a song; transcribe dialog of a TV episode; subtitle a movie	Enthusiasts	1
Create remix	Create a playlist; create a “mash-up” of two or more songs; create a music video	Remixers	1
Describe remix	Rating a music video; commenting on a mash-up; categorizing a playlist	Newbies	1
Find remix	Searching for a music video by video source; searching for a playlist by remixer; searching for a mash-up by song	Newbies	1

Use Case	Examples	Target Users	Priority
Request generation of remix	Specifying a song and video source to be used in a generated music video	Newbies	2
Browse ontology	Browsing structured descriptions of fictional characters	Enthusiasts	1
Search ontology	Find everything that has been asserted about a particular TV series	Enthusiasts	2
Contribute to ontology	Create a relation between a fictional character and an actor	Enthusiasts	1
Contribute to ontology schema	Create a new type of relation that can exist between characters	Enthusiasts	2
Summarize media	Request a 30-second summary of a feature-length film; catch up on what has happened in a TV series	Newbies	2
Promote media	Alert potential users of an archived film to its existence; inform fans of a musician about a new song	Original creators	1
Protect copyrighted media	Do not allow redistribution of original media without the copyright owner's consent	Original creators	2

Functionality requirements

Public API

MSMDX must have a stable, open API for adding and retrieving data, with the following qualities:

- Any functionality accessible through MSMDX software components should be available programmatically by directly accessing this API
- Whatever kinds of data that can be added to the system should be retrievable through a query API
- All MSMDX components should use this public API—there should be no “undocumented” protocols for accessing MSMDX data
- *API functionality:*
 - Add a reference to a media object (image, audio, video, timed text, composite playlist)
 - Add an annotation relationship between a media object and a media object fragment (former annotates the latter)
 - Resolve a reference to a media object or media object fragment, given a description of its content

- Retrieve all annotations of a media object
- Retrieve a specific user's annotations of a media object

Media player plugins

Plugins should be developed for major media players, beginning with Windows Media Player. Ideally the WMP plugin would be developed in such a way as to make it usable by other media players which are based on Windows Media components. These plugins should provide the following functionality:

- *Media consumption*
 - Handling of MSMDX playlist files
 - Display of associated metadata for media fragments presented in a playlist
 - Jumping from media fragments presented in a playlist to the source contexts from which those fragments were taken
 - Handling of timed-text files provided separate from the primary visual or aural media they are intended to accompany
- *Metadata production*
 - Frame-accurate selection of temporal regions for annotation
 - Selection of polygonal spatial regions (ideally aided by edge-detection techniques) for annotation
 - Assistive interface for suggesting possible media annotations
 - Creation and synchronization of timed-text files

Structured metadata plugins

The media player plugin described above will provide a way for MSMDX users to quickly add simple annotations for media, but for more structured metadata (such as additions to the domain ontology) a different approach is preferable. Most rich content metadata for media is currently being created in the form of HTML posted to online discussion forums, blogs, wikis, and other content management systems. Researchers have noted that this practice can be viewed as a form of unstructured annotation.³³

MSMDX tools for structured metadata production should seek to enhance this existing practice rather than replace it. These tools could integrate with major content management systems such as phpBB³⁴ to add functionality which allows users of these systems to easily add semantic markup to their HTML.³⁵ This semantic markup could then be harvested by MSMDX crawlers to add data to the metadata repository.

33 David R. Karger and Dennis Quan, "What Would It Mean to Blog on the Semantic Web?" in *Proceedings of the 3rd International Semantic Web Conference*, (Hiroshima, Japan: 2004).

34 phpBB is a particularly attractive choice as it is used extensively by fan communities for organizing their online discussions.

35 There are a few competing proposals currently under consideration for allowing this kind of metadata to be layered onto standard HTML. See for example Tantek Çelik and Adam Rifkin, "MicroFormats," <http://developers.technorati.com/wiki/MicroFormats>, and Mark Birbeck and Steven Pemberton, "RDF/A Syntax," <http://www.formsplayer.com/notes/rdf-a.html>.

Sequencing tool

While hypertext-focused tools such as phpBB have excellent potential for providing more structured metadata, ideally MSMDX users will produce metadata through hypermedia as well. MSMDX should provide a simple, free, open-source tool for sequencing and arranging pointers to media fragments, which can then be saved in an open format and shared with others. This tool should support adding both unstructured and simple structured metadata to media fragment pointers.

For example, an anime enthusiast might use the tool to create a guided tour of his favorite anime scenes, with textual commentary accompanying each scene. The commentary could be added as metadata for those scenes in the MSMDX repository. Since this metadata is being created as part of a work meant to be distributed to others, rather than as an exercise in trivia collection, it is likely to be of high quality.

Editor plugins

Finally, since skilled remixer are not likely to give up their tried and tested tools for MSMDX replacements, plugins are needed which will allow remixer to interact with the MSMDX API from the comfort of their favorite editing suites. At a minimum, this means the plugin must be able to query for media fragments and resolve the resulting pointers into a format the editing tool can handle. Ideally, the plugin will also be able to export finished remixes into the open MSMDX playlist format. Adobe Premiere should be the first editing tool targeted, as it seems to be the most widely used tool in the fan video remixing community.