Chapter Five: Discussion

5.1 Review of bittorrent website features

Of the thirteen sites looked at only two required registration to download torrents. Many sites limited uploading to registered members or in some cases specific groups of users. The site filemp3 is noteworthy in that it was much more structured and hierarchical in its users and their submissions. The classes of user and terminology used demonstrated a strong similarity to those described in previous studies of online piracy (Rehn 2001, Cooper and Harrison 2001) more so than other sites that indexed „pirate" material. Different users had different status and rights on the website and would obtain status through uploading a large amount of material, making a monetary donation to the site or through obtaining new („zero-day") releases.

All but one of the sites where users could give feedback on torrents required registration in order to do so. This could be to prevent „spam" being distributed in this way; xTVi was the only site allowing public feedback but did so in the form of flags and no free text field was available to leave user comments. Documentation on all of the sites was usually sparse, although if a subject had been dealt with by a relevant forum thread then it was often linked to. Where forums were available users were generally encouraged to use them for help and advice. Little site-specific documentation was available and that which existed was usually in the form of a very brief sentence or two on how to get started.

The range of guidance for uploading material and entering torrent information was very broad. Where it existed there seemed to be an emphasis on prohibiting types of torrents and then a reminder to include any available information in the torrent record. There was little evidence of moderation on the sites, although this does not mean it did not occur. The fact that moderation usually occurred after contributions had been made, together with the low barriers of entry to most sites would seem to indicate a lack „gate keeping" generally amongst the sites, where trust plays an integral part in building and maintaining the website data (Jones 2005).

The facility to give user feedback was usually in the form of free text comments only. Three of the larger sites did offer a means of rating torrents on a scale. The rating was a little ambiguous as to its specific meaning; it was not explicitly stated if it referred to the technical quality of the material or its semantic value. Isohunt was interesting in that it calculated its release rating on a combination of the two by using a metacritic[1] score in conjunction with user votes (based on technical quality). Although these rating systems did not seem to be a heavily used feature on any of these sites, it does give users additional information by which to differentiate between otherwise similar torrents.

The categories used by the sites were surprisingly inconsistent, with some sites using very broad terms and others using a greater number of smaller categories. With the exception of tlm-project, all of the sites had at most two levels of categories. All of the sites holding different media types such as audio or video would make this distinction at the top level, although the criteria for subcategories often seemed arbitrary. This might indicate that, especially in the larger sites, the structure does encourage easy browsing by users; studying the distribution of torrents among categories could possibly be used to verify this. As might have been expected, specialist sites used categories requiring a more advanced understanding of the context of the material by users.

The torrent sites allowed little opportunity for „chaining" i.e. movement between related records. When viewing a torrent record on the torrentportal website, similar records were displayed, based on matching terms occurring within the name fields. Other websites however, only allowed users to go back to a specific category or conduct a new search. Another possible exception to this might be isohunt, which was the only site to offer an alternative system of storing torrents. Isohunt has two levels of records – releases and torrents, with releases allowing multiple torrents to be linked to it. The responsibility of making the link between the two types of records is given to the user and presently the majority of torrents indexed by the site are harvested from other torrent sites and have remained unattached to release records. Although this approach may be not be wholly implemented on the site it is a clear recognition of the fact that electronic resources may be represented in different formats and locations as indicated by Heery (1996).

[1] www.metacritic.com is a website that records review scores for media from other sources such as newspapers and film websites and calculates an average score.

The actual torrent records themselves were quite simplistic – mostly comprising of name, description and category fields, with additional implicit metadata calculated by the server. This may be because of the wide range of media which can be used with bittorrent means that few elements of metadata would be relevant across this spectrum. It could also be that by adopting a standard format of name and description there is a higher degree of semantic interoperability (Heery, Powell and Day 1997) giving users a lower cognitive load when using different sites and also allowing records to be harvested and indexed by other sites with greater ease. Surprisingly, the specialist sites did not offer any additional elements that could be used to describe the material – artist or format fields in music torrents or platform in software torrents for instance. The private website demonoid, possessing the highest entry barrier for users, seemed to have the most detailed records by far; multiple additional fields, dependant on the torrent category, were mandatory in the creation of records by users.

Almost all of the sites looked at allowed searches to be conducted using query strings. In general there were relatively few advanced search features supported with little or no documentation for them. This suggests that users generally conduct simple queries, as has been found in studies of internet search engine users (Spink and Bateman 1998, Spink and Jansen 2004). An additional study of query terms used is needed to confirm this, however. Another fact supporting this is that often the default searches on the sites were as inclusive of records as possible, using OR as the default operator for multiple terms and applying wildcards automatically. In addition to standard browsing and searching, the busier sites offered RSS (Really Simple Syndication) feeds and sometimes email notifications that listed new torrents available through the website. This allows the activity of monitoring to be carried out more easily by users and can be customized to include different torrents according to their assigned category.

5.2 Content analysis of user-created metadata

The most striking result from the content analysis is how many torrent records used only the mandatory name and category fields and not the description field. Additional information in the description field was only used in 490 records – 52.4% of the total. Furthermore in over a

fifth of these the only additional information was a link to another torrent site, presumably where the material originated from. This means that 445 or 47.6% of the total number of torrents looked at described the material in the name field alone; a maximum of a hundred characters. In some cases this relatively small field contained many different metadata elements, for example the name field "[SKL][Strawberry Panic 01][DVDRiP][X264 AAC] mp4" consists of the following five elements (from left to right): release group, title, source, codec and file format. In other cases the name field will contain the Title element only, meaning the title will often only make sense in the context of the obligatory category field – one record was simply titled "Grudge" for example, although given that is was in the film category one might assume the material was the recent Hollywood movie „The Grudge".

Only 23.3% of the torrents contained a non-technical review of the material and around a third of these were two sentences or less. Compared to the occurrences of other elements in the sample this is quite frequent but perhaps surprising none the less. Links or references to other torrent websites occurred frequently particularly within the name field. The frequent presence of this element is interesting in that the value of its use for location or searching is apparently negligible (users searching for torrents from this site would be better visiting it first-hand) and its use in selection is arguable. It is a more likely possibility that this element is used to either respectively acknowledge the source of a torrent or to promote the site's use by the community. The release group element is not always applicable to torrents; individual users not affiliated with such groups contribute much of the material. As such, 24.5% is a relatively high rate of occurrence within the sample. While one could speculate that acknowledgement and promotion are again two likely reasons for its use, the element also has a direct bearing on the quality of the material.[2]

Among the elements specific to audio material, it was found that format featured most often as did the bitrate and genre. The high use of genre was surprising as the website offered 32 different subcategories based on genres, which torrents within the music category had to belong to. The source element was used less than bitrate although it also determines the quality of the material. When it was used however, it often denoted a source other than

[2] The relationship in the distribution of pirated releases could be likened to that of a restaurant; the original source material (movie recording, computer software) forms the ingredients, the release groups act as chefs in preparing it, before it is served by the waiters – the websites to the customer – end-user.

compact discs which might indicate that compact discs had become a standard source that was assumed and did not require stating.

In the video material torrents it was found that the source, file format and codec used were the elements occurring most frequently, by some margin. The source element was more prevalent than in the audio torrents which is likely due to the higher variance in its entries – the source in video material could be „regular" television, high-definition television, commercial DVD, or a number of different types of illegal recordings of cinema showings. Because these different types of sources all represent large proportions of the video material shared using bittorrent there is no easily identifiable standard as in the case of audio material.

The element most present in the software material was the version. This was partly expected as the version reference could well be considered part of the title field and, as a torrent may exist on a site for previous versions of the material, the version element allows users to distinguish between them. As with the genre element in the audio material, the platform element was often present although the information was already present in the choice of category used for the torrent. The name of the software publisher was very rarely used, perhaps because an external link to the software"s homepage was considered more useful.

Between all of the different sets of torrents the type of metadata that occurred most frequently was technical information affecting the compatibility of the material. This type of metadata includes the format, file format, codec, platform and version. All of these elements will determine if material will work on a user"s computer system and related hardware (such as portable devices and home entertainment systems). In the case of audio and video material, metadata that was related to the quality of the recording also occurred very often: the source, video specifications and bitrate elements. For software torrents, this is irrelevant as the material will be a straight copy of an existing digital source.[3]

[3] Audio and video torrents *may* be a straight copy from a digital source (CDs and DVDs respectively) but this is not necessarily the norm and as such is often acknowledged through the source and format elements of the metadata.

Of the torrents that were remaining after one week, only 11% had user comments attached to their records. Roughly two-thirds of these comments did not relate directly to evaluating the material and were used to otherwise communicate between users who were interested in the torrent. These comments were often expressing gratitude to the user who originally provided the material, requesting more users to seed or requesting and submitting advice on how to use the material.

Those comments that were evaluative were mostly negative feedback of a technical nature, relating to the material"s inability to work or, where relevant, the poor quality of the recording. This would indicate that users are more likely to leave a comment when a torrent"s material fails to work adequately, is described incorrectly (including the presence of a virus or spyware) or is of a lower quality than expected, as opposed to when a torrent"s material performs as expected. Users regard „genuine" torrents as the norm and therefore are less inclined to leave comments that simply confirm this. The lower proportion of non-technical feedback also supports the findings of the metadata coding; that technical information is regarded as more important to bittorrent users than non-technical details of the material being shared.

5.3 User questionnaire results

Respondents" answers to questions one and two revealed that they were mostly experienced users of bittorrent that used it for over a year and tended to use it often. Question three also determined that around half of the respondents also used other p2p networks, perhaps finding them more suitable for locating certain types of material.

The responses to question four show that bittorrent websites are by far the most popular method of finding torrent files. Some bias is possible given that bittorrent website forums were one of the primary places where the questionnaire was promoted but the alternative methods of locating torrents are still very low – only 7% of respondents used regular search engines more than occasionally and the corresponding percentage for user-sent torrents was only 3.2%. The most popular method was through using multi-category sites, although specialized and sites that indexed others were also very popular. These meta-searching

sites seemed to be less of a first choice for people than other bittorrent sites but many more respondents used them „not often", „occasionally" or „quite often" which might indicate that they constituted a second choice for many users not able to find material on their usual sites. Given the wide availability of RSS feeds and emailing lists on torrent websites there seems a relatively low take-up by users in monitoring new information through this method.

The optional response to list an alternative method was used by few respondents and largely consisted of more specific sites or types of sites that were already listed – particularly private sites, which would be either be multi-category or specialized sites that required registration or invitation. Three different methods of finding torrents not listed in the choices were identified – forum communities (three respondents), blogs (one respondent) and torrent harvester, a program that performs meta-searches by crawling several different torrent websites to match query terms (one respondent).

The responses to question five were not what might be expected, given the previous results which suggested bittorrent websites favoured searching rather than browsing in locating records. Where the information need was not fully defined (as with the first to third parts of the question) semi-directed browsing (Marchionini 1995) was the most popular way to locate torrents, closely followed by performing searches. Undirected browsing (Ibid.) was also surprisingly popular – one might have thought that given the additional time and resources involved in downloading bittorrent material – as opposed to web „surfing" – users might be less likely to find material through chance that they would be willing to retrieve. Almost half the respondents used the sites to search for known torrents „very frequently". Part of this is due to the efficiency with which piracy takes place; users will expect that newly released material will be distributed through bittorrent and will often know who does this (the individual or release group) and where it will be found. It should be noted that while this question provides an idea of the means by which the respondents locate torrents, more research in this area would be required to produce a more detailed analysis of their behaviour.

The results of question six indicated a high level of involvement with the community – over half the respondents participated in some way towards the websites material, and a fifth of respondents contributed new material or torrent files. As with question four there is likely to be some bias in the respondents who answered because of its distribution; in this case the

level of participation in forum discussions may be non-representative of bittorrent users on the whole.

Questions seven to ten provided user ratings of metadata elements, many of which were directly comparable to those looked at in the content analysis (see appendix four). Generally speaking, there was a loose correlation between the mean average of the ratings given by the respondents and their frequency in the content analysis torrents; the elements that were rated most highly by users were also the ones most often present. There were exceptions to this, however; for example the platform and language elements in the software torrents were the most highly rated in the questionnaire but were second and fourth in frequency. In the case of the platform element, the information was already available through the category where the torrent was placed and hence effectively redundant when in the name or description field. With the language element, in all of the types of torrents, there is perhaps an assumption made by users, similar to the source element for audio torrents. Users may assume that, because the website is written in English and the torrent"s title is in English, then the material is the same, unless otherwise stated.

Metadata relating to the technical quality (source, codec, bitrate and format) were rated highly in both the audio and video categories. Among other, non-technical metadata there was often a high degree of variance, with large numbers of users rating elements at both ends of the scale. External links and reviews, by third parties or by the person uploading the material, were usually given a relatively low score by respondents. The responses indicated that users were generally quite confident in understanding the metadata types as across the different categories there was a very low occurrence of „unsure of" answers. The responses to question ten – metadata applicable to all torrents – indicated that users were generally more concerned with the availability and speed with which material could be obtained than where it was located, the date it had been uploaded or how many times it had been downloaded. User comments were also rated quite highly.

Material in the literature review had covered release groups in the context of software piracy and respondents were therefore asked about this metadata element in question seven (for software torrents) only. When the questionnaire was made available to respondents, the content analysis had not yet been fully completed. As a result of this, the extent to which

metadata relating to release groups was present in all categories of the torrents was not anticipated. During the content analysis it became apparent that release group metadata was also prevalent in audio and video torrents. Despite respondents not being able to rate this element for these two types of torrents, five respondents mentioned it specifically in the free-text part of question ten, which would suggest that release group metadata is of some significance to many bittorrent users.

In order to see if there was a variation in the rating of metadata elements between those who uploaded torrents and those who did not, the results of questions seven to ten were cross-tabulated with users who indicated (in question six) that they created torrent material or new torrents (see appendix four). Of the two groups of users those who created torrent records rated the technical metadata elements more highly but this was not by a significant degree – on the whole there was a general consensus in how important the different metadata elements were.

The responses to the first three parts of question eleven failed to distinguish any prevailing trends in user opinions. There were large proportions of the respondents both agreeing and disagreeing that torrents often lacked information, although few felt strongly about this. Similar results were obtained on the issues of accuracy of information and irrelevant information. The last part did reveal that only a small section of respondents (18.1%) had difficulty with the terminology used in torrent records. This was surprising as this was a particular difficulty encountered in conducting the content analysis; there seemed not to be a single set of identifiable abbreviations and terminology used in the torrent records but rather many different communities each using different variations.

Question twelve asked users how often they encountered difficulty in locating torrents that met their needs. The result was that only a small fraction (4.7%) of the respondents „often" or „generally" had difficulty while the majority of users (74.4%) either „rarely" or „almost never" had difficulty. This was again a surprising result given that in the previous question many users did not express a strong positive opinion of the torrent metadata. Because of this it appears that, while most of the respondents acknowledged that there is room for improvement in the information given about torrents, it does fulfil its function in enabling users to find and evaluate the material available.

The results of questions eleven and twelve were cross-tabulated with those from question one in order to see if experienced users of bittorrent were significantly more successful in locating material than newer users. The results (see appendix four) did show a slight trend with newer users, as might be expected, having marginally more difficulty with the terminology used and also in successfully finding material.

The final question allowed users to contribute feedback regarding the study. Of those comments left, several referred to private bittorrent sites as being a better source of torrent information and in terms of quality of material than public torrent websites. Some of the answers to question four also indicated that many users distinguished between the two types of website. It is interesting that private websites, which have a higher barrier to entry and sometimes a larger amount of gate-keeping for torrents submitted by users, should be identified as providing a better service. The two private websites looked at in this study were more structured and appeared to have a more active community; evident from forum posts and also user comments. One might speculate that it is one, or both, of these two factors that positively affects the performance of these sites: more structured gate-keeping and greater community participation.