

Ofcom
Illegal Filesharing Pilot -
Peer Review

Prepared for:
Ofcom

Agency Contacts:
Dave Chilvers

Prepared by:
BDRC-Continental
Kingsbourne House, 229-231 High Holborn,
London WC1V 7DA
t: 020 7490 5944
f: 020 7490 1174

Month of Report:
August 2010

Job Number:
J10168

Title	Page
1. Background.....	3
2. Objectives	4
3. Summary of recommendations	5
4. Detailed assessment of pilot study	8
4.1. Recommended methodology	8
4.2. Data collection	9
4.3. Questionnaire – generic construction.....	9
4.4. Questionnaire – specific questions	10
5. Specific results and analyses	12

Appendix – Pilot questionnaire

1. Background

Under the Digital Economy Act 2010, the remit of Ofcom has been extended to include a range of new duties related to online copyright infringement. As part of this work, Ofcom is required to assess the level of use of the internet to infringe copyright. In order to achieve this, Ofcom needs to develop a consistent and representative measurement system that is able to gauge the level of activity and attitudes towards illegal file-sharing in the UK over time.

In March 2010, Kantar Media conducted a pilot survey on measuring the extent of illegal file sharing and downloading. Telephone, online and face to face methodologies were used among adults and children aged 12-15 (mainly omnibus surveys given timescales and cost). The pilot research aimed:

1. To compare quantitative data from equivalent online, face-to-face and telephone surveys - to identify whether any significant differences exist
2. To explore consumer understanding and interpretation of the questionnaire content
3. To establish which methodology yields the most accurate response data – particularly in relation to the perceived difficulty of obtaining honest responses in areas of potentially (and actual) illegal activity
4. To explore whether file-sharers are more or less likely to participate in any particular survey methodology, whether the profile of those participants vary, and whether there is a difference in the responses they give

A written report was produced on the research itself which also highlighted recommendations as to how this difficult area could be monitored moving forward.

2. Objectives

The main objective of the peer review is to identify any weaknesses with the proposed measurement approach, particularly any that could have a significant impact on the survey findings, so that these can be considered and a decision made as to whether the proposed methodology needs to be amended in any way. The intent is to ensure that the approach is as robust as it possibly can be, and that any limitations of the chosen methodology have been fully evaluated. In that way, Ofcom can be fully prepared for any challenges to its approach.

To undertake the review, we have been granted access to:

- The questionnaires used in the pilot survey
- The report produced by Kantar
- The tables generated from the various pilot surveys
- Respondent level data from all the pilot surveys

In providing our assessment of the proposed methodology, we have used our general experience of undertaking research to produce an appraisal of the recommendations made, and in so doing have undertaken some further analysis of the respondent level data.

In this report we set out a summary of recommendations (in Section 3 immediately following) and more detailed analysis which supports these recommendations (in Section 4 et al).

The key considerations, which the Kantar report on the pilot acknowledges are:

- **Representativeness** – ensuring the proposed data collection methodology and sample design properly cover the target market
- **Honesty** – ensuring the questions deliver accurate data and that respondents are not likely to amend their responses for fear of admitting to illegal activity

3. Summary of recommendations

Recommendation 1:

The proposed methodology, combining online and face to face data collection, is the best approach to ensure both high and low frequency Internet users are covered in the tracking research.

Recommendation 2:

Demographics and frequency of internet use should be used to rim weight the tracker data. The latter will ensure that frequent internet users (with greater likelihood of downloading or sharing files) are weighted correctly in the overall dataset. The profile of internet usage can be taken from screening questions for the face to face part of the tracker.

Recommendation 3:

Implied levels of illegal activity are likely to be more accurate than admitted levels of illegal downloading and file sharing. As such, the question used in the pilot which covers types of website visited (peer-to-peer, commercial, social networking etc) is probably the best way to obtain accurate information.

However, this question needs to be asked separately for each category of downloaded item (music, software, video games, films etc) rather than just being asked once as in the pilot (and this may well have been for time and cost reasons). In this way, illegal activity for each category can be more accurately assessed (e.g. for most listed categories, it is likely that only commercial websites represent legal activity).

In particular, the way the question was asked in the pilot may have included types of files shared or downloaded in addition to those listed in the preceding question (music, software, video games, films etc) for which sharing is perfectly legal.

Recommendation 4:

Consideration needs to be given in the tracker as to how to best disguise the desire to measure illegal activity. In the pilot, this was possible as the questionnaire was just 5-7 minutes and part of a more general omnibus questionnaire. For a 25 minute tracker questionnaire, inclusion in an omnibus is unlikely to be feasible.

The tracker therefore needs to include some more general questions upfront about Internet usage and the question about types of website visited and frequencies needs to be as near the start of the questionnaire as possible.

A more general point is that whilst a question about use of different types of website in each category is likely to yield more honest responses than a direct question about illegal activity, there is still a possibility that illegal downloaders will not admit to using e.g. peer-to-peer websites. One option in the tracker – given it will be a much longer questionnaire – could be to ask an open ended question about which websites have been visited in the last week – possibly asking the respondent to nominate up to 10 websites they visited during that period. If this appears early on in the questionnaire as part of a section about general Internet use, respondents may mention sites which allow them to be classified as illegal downloaders (and which the more direct question on whether peer-to-peer sites have been used may miss).

Recommendation 5:

A methodological alternative worth consideration is to use consumer omnibus surveys (face to face and online for the reasons given earlier to capture both low and high frequency Internet users) for key questions and use a parallel ad hoc survey to pick up detailed data. The use of the omnibus approach will help camouflage the focus of the questions but it will be necessary to augment the omnibus questions as outlined in section 4 below. A decision can be made on this once the key questions have been finalised and the feasibility of including this questionnaire set on consumer omnibus surveys determined (the key issue will be the length of the questionnaire containing these key questions – see section 4.4 for our recommendations).

Recommendation 6:

Commercial web tracking monitors such as UKOM from Nielsen should be evaluated to assess whether downloads from individual peer-to-peer sites and aggregate data from the top such sites can be measured on a regular basis. This information could provide the hard universe data on top of which the proposed tracker could provide underlying attitudes and behaviour. It could also be used as a rim weight for the tracker in addition to demographic and frequency of internet use rim weights.

Recommendation 7:

Surveys of children aged 12-15 should be included if Ofcom wishes to obtain a more complete picture of illegal file sharing and downloading. However, in conducting these surveys, two points need to be borne in mind:

- The decision to use 12 as the minimum age could be seen as arbitrary and there may be file sharing and downloading activity among younger children
- The inclusion of surveys among children will likely add disproportionate cost, given that these will need to be ad hoc surveys – even if a solus online option is taken (and this could lead to challenge that it over represents heavy internet users as per the adult surveys and subsequent recommendation for a mixed mode survey)

4. Detailed assessment of pilot study

4.1. Recommended methodology

In an ideal world, online research would be the obvious way to conduct this tracker. However, the pilot surveys replicated previously well known conclusions that members of online panels (from whom the online omnibus was selected) use the Internet to a much higher frequency than Internet users in general (93% online respondents use the Internet at least once a day compared to 76%/74% from the other two methods of data collection). It is not even possible to upweight the low frequency Internet users from the online survey as there are so few of them.

High frequency Internet users have higher usage of download services and so without correction a solus online methodology would bias the results (across all three methodologies, 50% of those using the Internet at least once a day admit to any file downloading/sharing, compared to 17% for those using the Internet 1-6 times a week and 5% for those using the Internet less than once a week).

A mixed methodology is therefore the only practical approach, with an online survey providing data for frequent online users and a face to face survey providing data for less frequent users (and as a by product of the screening process – assuming an overall representative sample – providing data to weight the two groups together utilising frequency of Internet usage).

However, the suggested questionnaire length for the tracker (25 minutes) makes it unlikely that the omnibus approach used in the pilot would be feasible in practice. And even if it were, the advantage the pilot conferred of including a short section on (potential and actual) illegal activity within a survey about other subjects would be lost if the majority of the questionnaire were about this illegal activity.

It is therefore very likely that an ad hoc survey will be required to administer the tracker and this therefore requires further consideration of how to minimise the likelihood of respondents failing to admit to activities that they might/would see as illegal (more detail is provided about this under Section 4.3 – questionnaire construction).

4.2. Data collection

The pilot produced evidence that data collected without the involvement of an interviewer was likely to produce more honest responses than that provided in response to interviewer directed questions.

The pilot report recommends that for the face to face element of the proposed tracker, a self completion methodology is used to collect the data about file sharing. For CAPI, this would involve the respondent using the laptop on which the questionnaire resides, which given that all such respondents would be reasonably proficient with a keyboard, should not present a problem. It is possible that the face to face element could be administered via a paper questionnaire and again this affords the option for the respondent to be given a form to complete to cover this part of the questionnaire. Indeed, this option would perhaps be more comfortable for the less frequent Internet users at whom the face to face part of the tracker would be targeted.

This approach should work in practice, as long as the self completion element is not the bulk of the questionnaire. If this is not the case, the interview is likely to become quite onerous for the respondent, as he/she will be spending the majority of the time both reading and answering large numbers of questions (rather than just answering questions).

4.3. Questionnaire – generic construction

Given that the tracker is very likely to be an ad hoc survey, it is important that steps are taken to encourage **honest** responses to questions where respondents may be admitting – either directly or indirectly – to illegal activity.

To minimise the impact of this, we recommend the following:

- Some general questions about technology used in the home in general and Internet behaviour specifically are included at the start of the questionnaire
- Inclusion of an open ended question regarding websites visited recently. If this appears before any questions about peer-to-peer group activity, it may highlight the use of such sites without prompting

- The question on types of website visited (Q4 in the pilot) is asked after these general questions and is asked for each category of download/file sharing
- Any questions about illegal activity should be placed at the end of the interview. The aim should be to deduce the incidence and volume of illegal activity rather than ask it directly
- However, the attitude questions should be placed after questions about behaviour and specifically about illegal downloading/file sharing directly. If attitude questions are asked before these behavioural questions, it is possible that there might be some influence. For example to be asked to agree or disagree with the statement “It is easy to find files on the internet for free that would usually be paid for” could be seen to confer some respectability onto free (and possibly illegal) downloads as opposed to paid for (and probably legal) downloads and this might affect how much respondents admit to illegal downloading

4.4. Questionnaire – specific questions

To deduce illegal activity, we recommend the following approach for each category of downloading/file sharing (music, video, software etc) following the existing Q2/3 which should be split to cover downloading and file sharing separately:

- New Q3x: Please type in/tell me the names of websites you have visited in the last week. You can give me up to 10 websites but try to select those you have visited the most frequently in the past week
- Existing Q4: Which of these have you used in the last 4 weeks in order to download **music** files through the Internet? [responses as per existing Q4]
- New Q5a1: [If used commercial websites] How many music tracks do you think you have downloaded in the past 4 weeks from commercial websites – please count an album as 10 songs. [This question can be used to assess legal downloads/file shares]

New Q5a2: [If used any other sites] How many music tracks do you think you have downloaded in the past 4 weeks from other websites – please count an album as 10 songs. [This question can be used to make a rough assessment of illegal downloads/file shares, as some use of peer-to-peer and other non commercial sources may well be legal]

New Q5a3 [If Q5a2 > 0] And what percentage of these tracks did you pay for?

Repeat for those file sharing music.

Repeat for other categories.

By asking Q4 for each category, we should avoid the situation which probably arose in the pilot whereby respondents included genuinely legal types of file download from sites other than commercial websites.

It will be essential to try and minimise number of categories to alleviate respondent boredom – it would be worth considering the merging of downloading and filesharing into a composite question if there is no great value in obtaining separate data for each. This will be particularly important if these key questions are placed on omnibus surveys, with a parallel ad hoc survey undertaken to provide more detail.

5. Specific results and analyses

Given that the focus of the tracker will be to measure illegal activity, we have analysed the pilot data to help assess this. We have developed three definitions of an “illegal” downloader:

1. A respondent using any of the file sharing/downloading methods at Q4 other than “commercial websites” - we have called these deduced illegal downloaders. Not all of these downloads will be illegal as the respondent could have assigned downloading pictures from e.g. Facebook in this category in the pilot survey (we have recommended how to improve this for the tracker)
2. Only allowing this definition if the respondent does download or file share one of the first six categories shown at Q3 (music, software, video games, films, TV programmes, books). This is likely to further eliminate legal downloads and come closer to a true measure of illegal activity
3. A respondent where less than 100% of material downloaded was not obtained legally for one or more of the six categories at Q9 – we have called these admitted illegal downloaders

Using these definitions, the percentages of the **total** weighted sample by gender and age are as follows:

Subgroup	Deduced illegal downloaders	Deduced illegal downloaders – filtered to only include those downloading 1+ relevant category	Admitted illegal downloaders
Total	33.9%	30.9%	17.1%
Male	38.1%	35.0%	20.4%
Female	29.7%	27.0%	13.9%
12-15	48.7%	43.4%	30.1%
16-24	50.7%	47.6%	31.6%
25-34	38.9%	35.2%	19.7%
35-44	36.3%	33.7%	16.2%
45-54	27.8%	25.5%	10.8%
55-64	21.6%	19.4%	6.1%
65+	5.6%	5.1%	1.4%

The pattern of illegal downloaders is very similar for the three definitions (males and younger respondents higher) but the percentages are far higher using the “type of file sharing site used” than by using the question where some illegal downloading is admitted. This reinforces the dangers of using an explicit definition to measure illegal downloading rather than an implicit one. If the deduced illegal download figure is correct, it is twice the size of the estimate produced by asking the direct question.

Of course, those responding to Q4 (methods used to share or download files) may have had things other than the categories in Q3 in mind when talking about file sharing (e.g. photos they had taken) and the question will clearly need tightening in the tracking survey – even when filtering on those downloading or file sharing one of the six categories. This reinforces the recommendation (Recommendation 3 in Section 3) that Q4 be asked for each category of download.

On the other hand, some illegal downloaders may have been reticent to admit to using peer-to-peer sites and this observation has led us to the recommendation that the names of individual websites visited recently and frequently should be captured via an open ended question and coded subsequently into categories (including those listed at the existing Q4).

Appendix – Pilot questionnaire

ASK ALL

QA: How many mobile phones in total do you and members of your household use?

Single-code

1. One
2. Two
3. Three
4. Four or more
5. None
6. Don't know

ASK ALL WITH INTERNET ACCESS

QB: Thinking of all the things you use the Internet for, and regardless of where and how you access it, please tell me overall how often you personally access the Internet?

Single-code

1. Several times a day
2. Once a day
3. 5-6 days a week
4. 2-4 days a week
5. Once a week
6. 2-3 times a month
7. Once a month
8. Less often
9. Don't Know

ASK All

Q1: Which of the following do you think the internet can be used for?

Multi-code

1. Make telephone calls
2. Read the news
3. Download and share music, video games, software, films, TV programmes and video
4. Watch live TV programmes
5. Look at maps from around the world
6. Book train tickets
7. Vote in a general election

ASK IF Q1 = 3

Q2: Have you ever downloaded or shared any of the following through the internet? If so, which ones?

Multi-code, Rotate

1. Yes – music
2. Yes – video games
3. Yes – software or applications
4. Yes – films
5. Yes – TV programmes
6. Yes – books
7. Yes - podcasts
8. Yes – other type of file. Please specify
9. No
10. DK

ASK IF Q2 = 1 to 7.

Q3: Have you downloaded or shared any of the following through the internet in the last three months? If so, which ones?

Multi-code, Show in same order as Q2

Only show codes mentioned at Q2

1. Yes – music
2. Yes – software or applications
3. Yes – video games
4. Yes – films
5. Yes – TV programmes
6. Yes – books
7. Yes - podcasts
8. Yes – other type of file. Please specify
9. No
10. DK

ASK IF Q2 = 1 to 7

Q4: Which of these have you used in the last three months in order to download or share files through the internet?

Multi-code, Do not rotate

1. Peer-to-peer such as BitTorrent, Gnutella, eDonkey, Limewire and Ares
2. Commercial websites such as iTunes, Blinkbox, Amazon, Lovefilm, Movieflix, Napster, Play, or Spotify
3. Social Networking sites such as Facebook, Myspace, or Bebo
4. File sharing websites such as Rapidshare, Yousendit, or Easyshare
5. Messaging programs such as Windows Messenger or Skype
6. FTP
7. Email
8. Other (please specify)
9. Can't remember

IF Q3 = 1

Q5a1: How many music tracks do you think you have downloaded through the internet in the last 4 weeks? Please count an album as the equivalent of ten songs

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1 to 5
3. 6 to 10
4. 11 to 20
5. 21 to 50
6. 51 to 100
7. Over 100
8. DK
9. Refused

IF Q3 = 1

Q5a2: How many music tracks do you think have you shared with others through the internet in the last 4 weeks?"

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1 to 5
3. 6 to 10
4. 11 to 20
5. 21 to 50
6. 51 to 100
7. Over 100
8. DK
9. Refused

IF Q3 = 2

Q5b: How many video games (excluding patches and upgrades) do you think you have downloaded through the internet in the last 4 weeks?"

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1
3. 2
4. 3
5. 4 to 10
6. 11 to 20
7. 21 to 30
8. 31 to 50
9. Over 50
10. DK
11. Refused

IF Q3 = 3

Q5c: How many software products or applications (excluding patches and upgrades) do you think you have downloaded through the internet in the last 4 weeks?"

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1
3. 2
4. 3
5. 4 to 10
6. 11 to 20
7. 21 to 30
8. 31 to 50
9. Over 50
10. DK
11. Refused

IF Q3 = 4

Q5d1: How many films do you think you have you downloaded through the internet in the last 4 weeks?

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1
3. 2
4. 3
5. 4 to 10
6. 11 to 20
7. 21 to 30
8. 31 to 50
9. Over 50
10. DK
11. Refused

IF Q3 = 4

Q5d2: How many films do you think you have you shared with others through the internet in the last 4 weeks?

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1
3. 2
4. 3
5. 4 to 10
6. 11 to 20
7. 21 to 30
8. 31 to 50
9. Over 50
10. DK
11. Refused

IF Q3 = 5

Q5e: How many TV programmes do you think you have you downloaded through the internet in the last 4 weeks? Please do not include the use of iPlayer, ITV Player, Sky Online or 4OD

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1
3. 2
4. 3
5. 4 to 10
6. 11 to 20
7. 21 to 30
8. 31 to 50
9. Over 50
10. DK
11. Refused

IF Q3 = 6

Q5e: How many books do you think you have you downloaded through the internet in the last 4 weeks?

ALLOW RESPONDENT TO CODE ACTUAL NUMBER AND CODE INTO BANDS BELOW

IF RESPONDENT STATES DK ASK THE FOLLOWING BANDS DIRECTLY:

1. None
2. 1 to 5
3. 6 to 10
4. 11 to 20
5. 21 to 50
6. 51 to 100
7. Over 100
8. DK
9. Refused

ASK ALL

Q6. Which of the following, if any, do you think can be illegal to download?

1. Music
2. Software
3. Video games
4. Films
5. TV programmes
6. Books
7. None of these

ASK ALL

Q7. How much do you agree or disagree with the following statements?

- a) I think that you should be able to download what you want from the internet for free
- b) It is easy to find files on the internet for free that would usually be paid for
- c) I often download free files on the internet to try them out before I buy them
- d) I find music and video files that you pay for on the internet expensive

1. Strongly agree
2. Tend to agree
3. Neither agree nor disagree
4. Tend to disagree
5. Strongly disagree
6. DK

ASK Q8-Q9 IF Q3 = 1 to 6

REPEAT Q8 AND Q9 IN TURN FOR EACH TYPE DOWNLOADED/SHARED AT Q3.

ROTATE ORDER

Q8. Thinking about the <category from Q3> you have downloaded or shared through the internet in the past three months, approximately what proportion of this did you pay for?

1. All of it
2. Over three quarters
3. Between half and three quarters
4. Between a quarter and a half
5. Less than a quarter
6. None
7. DK

Q9. And approximately what proportion would you say you obtained legally?

1. All of it
2. Over three quarters
3. Between half and three quarters
4. Between a quarter and a half
5. Less than a quarter
6. None

DK



BDRC-Continental, Kingsbourne House, 229-231
High Holborn, London WC1V 7DA

t: 020 7400 1000
f: 020 7490 1174