

YouTube comments as media heritage

Acquisition, preservation and use cases
for YouTube comments as media heritage records
at The Netherlands Institute for Sound and Vision

Archival studies (UvA) internship report by Jack O'Carroll



Contents

Introduction	4
Overview	4
Research question	4
Methods	4
Approach	5
Scope	5
Significance of this project	6
Chapter 1: Background	7
The Netherlands Institute for Sound and Vision	7
Web video collection at Sound and Vision	8
YouTube	9
YouTube comments	9
Comments as archival records	10
Chapter 2: Comments as audience reception	12
Audience reception theory	12
Literature review: Audience reception and social media	13
Conclusion	15
Chapter 3: Acquisition of comments via the YouTube API	16
YouTube's Data API	16
Acquisition of comments via the YouTube API	17
YouTube API quotas	17
Calculating quota for full web video collection	18
Updating comments collection	19
Distributed archiving with YouTube API case study	19
Collecting 1.4 billion YouTube annotations	19
Conclusions	20
Chapter 4: YouTube comments within FRBR-style Sound and Vision information model	21
FRBR at Sound and Vision	21
YouTube comments	25
YouTube comments as derivative and aggregate works	25
Alternative approaches	26
Option 1: Collect comments and treat them as analogue for the time being	26
Option 2: CLARIAH Media Suite	27
Option 3: Host using an open third party	28
Chapter 5: Discussion	29
Conclusions summary	29
Discussion: Issue of use cases	29
Possible use cases	30
Audience reception use case	30

Historical context for a video	31
Community or participatory archive use case	31
Crowdsourced metadata	32
Alternative perspective: YouTube Creators are primary users of comments	32
YouTube Creators are the new “media professionals”	33
Chapter 6: Legal considerations	34
Copyright and access	34
YouTube Terms of Service	35
Improving the YouTube archive	35
Chapter 7: Recommendations	36
Works Cited	37

YouTube comments as media heritage

Acquisition, preservation and appraisal for YouTube comments as media heritage records at The Netherlands Institute for Sound and Vision

Internship report by Jack O'Carroll, Archival Studies student at the University of Amsterdam, 2019

Introduction

Overview

This research project is about approaches to using YouTube and YouTube comments at The Netherlands Institute for Sound and Vision. The research will develop approaches to using YouTube comments that fall within the strategic goals and activities of Sound and Vision. The first chapter establishes the background of the project with regard to Sound and Vision, YouTube and archival theory. The second chapter considers comments as records of audience reception and looks at related research. The third chapter looks at the acquisition of comments via the YouTube API and the opportunities and restrictions that method presents. In the fourth chapter the project will consider how comments would fit within the framework used at Sound and Vision, as well as suggesting alternative approaches. The fifth chapter is a summary of concluding remarks and a broader discussion of the project, strategic goals, use cases and users at Sound and Vision, followed by legal considerations for collecting YouTube comments. Finally the report will give recommendations in terms of the technical and practical side of acquiring YouTube comments as well as suggesting ideas about strategic positioning of Sound and Vision in relation to YouTube and YouTube creators.

Research question

The question behind this project is:

How can YouTube comments be utilised by Sound and Vision as a media heritage archive?

Methods

Methods used for this project were a mixture of internal interviews, analysis of comments, literature review, an applied example and case studies. This mixture of methods was chosen because the project is very open-ended and exploratory rather than having a defined goal.

Internal interviews were conducted to get an understanding of the organisational goals at Sound and Vision, current and past projects relating to YouTube, and the strategic vision with regards to the broadening of scope in terms of collecting and presenting media heritage beyond broadcasting (for example web video and video games). Interviews also generated potential use cases for YouTube comments at Sound and Vision, including as records of “audience reception”.

The literature review in Chapter 2 was used to get a broad view of research and archival-type projects relating to social media and comments that could be said to treat these as audience reception records.

Approach

The approach of the research was open-ended and exploratory because this is a new area of interest that comes from a shift in strategy. It was difficult to create a more focused and precise research project relating to collecting YouTube comments because it depends on use cases, which depend on user research and more focused questions.

Secondary questions raised by the primary research question that this project explored were:

- What is a YouTube comment? What role does commenting play on the platform and what is the significance of a comment collection for a media heritage archive?
- What are the strategic goals and ambitions of Sound and Vision with regards to YouTube, comments and social media?
- What is the history of Sound and Vision and how does this project fit with existing projects and their primary work as a broadcast archive?
- What is YouTube and how does archiving YouTube differ from public broadcasting?

Scope

There were limitations placed on the scope of the project. For example, to experiment with the capabilities of the YouTube API a certain level of programming expertise is required to write scripts that will allow for systematic archiving within the API request quota. In terms of developing a concrete use case for comments this was also only possible to a limited extent and this led to further limitations on what type of appraisal and information model would be used due to the lack of real user research on which to base a use case. Another limit on the scope was the researcher being limited to the English language rather than Dutch and also being unfamiliar with Dutch YouTube and Dutch media. Sound and Vision is a Dutch broadcasting archive and heritage organisation.

Significance of this project

This study is unique as there are few examples of organisations archiving YouTube as cultural heritage in a systematic way. Furthermore within the field of archiving approaches to archiving user-generated social media features like YouTube comments is not something that is widely discussed. There is a growing interest in collecting user-generated data like social media and comments among academic researchers.

Chapter 1: Background

This chapter outlines the context of the project by giving background information on Sound and Vision, YouTube, comments and archival practice.

The Netherlands Institute for Sound and Vision

Sound and Vision is primarily a Dutch broadcasting archive. It is one of the largest audiovisual archives in Europe with over 1 million hours of radio, television, film and music. It was established in 1995 by the merging of the Amsterdam-based Stichting Film en Wetenschap's collection, the film archive of the Rijksvoorlichtingsdienst and the Broadcasting Museum. It is now the de-facto Dutch national audiovisual archive and specialises in digital preservation of audiovisual material. The primary users of the archive itself are media professionals for the purposes of re-use or research.

This primary function of Sound and Vision is mandated by Dutch Media Law and agreements are in place with broadcasters and production companies to carry out this work. The institute is also situated in Hilversum Media Park, a centralised location for broadcasting in the Netherlands where the offices of the NPO (the administrative body for Dutch broadcasting) can also be found as well as most of the main broadcasters themselves and various related service businesses. This relationship between Sound and Vision and broadcasters has been even better improved by the development of systems for acquisition and access of material such as establishing an automated digital process for acquisition of broadcast output, as well as DAAN, a media management platform that forms the basis of the core archive system, which is optimised for use by media professionals. However, his primary function is not the only public task that Sound and Vision is responsible for.

As well as being a broadcast archive it is a museum and knowledge institute. The museum has both permanent and temporary exhibitions that focus on Dutch media heritage and utilising archival materials to be presented to the general public. In 2020 the permanent museum is being re-designed and moving away from a historical broadcast museum to a more dynamic media experience (Middag). This is part of a wider strategic shift and broadening of scope for the organisation from being focused mainly on broadcasters to being interested in a wider media landscape that includes things like YouTube and social media.

As a knowledge institute Sound and Vision works with educational institutions, academic researchers and has its own research and development team. Within this part of the organisation that more experimental and innovative approaches to media and archiving such as the Dutch video game

archive and exhibition pilot in 2018. This is also where this project fits within the organisation as it seeks to broaden the scope of collecting and presenting material to include things like social media.

YouTube comments offer an interesting opportunity in this context. Comments on YouTube relate to audiovisual material (Sound and Vision's area of expertise), they are (potentially) very accessible in bulk via the YouTube API, and they are an opportunity to collect a new type of user-generated social media. Sound and Vision also already collects YouTube videos as part of a web video collection, so comments could be complementary to that collection.

Web video collection at Sound and Vision

Sound and Vision's web video collection began in 2007 as part of a research project - just two years after YouTube first started. Since then they have collected 10,000 web videos from platforms like YouTube but also from early web video platforms that no longer exist. Over time, the processes for collecting web videos have changed and become more systematic and procedural. For example, Sound and Vision contacts creators or intermediaries to get them to sign a license allowing the archiving of their material so as to avoid infringing rights such as intellectual property rights by archiving the content.

Martijn van der Vliet works on this project and is involved with contacting creators says that the process is still time-consuming and difficult (van der Vliet). Problems stem from the fact that it is difficult to contact creators of web video and sometimes the more professional creators can only be contacted through an intermediary company who they hire to deal with licensing and issues like people stealing their material. These companies are usually motivated by profit and can often refuse to allow videos to be archived. The strategy is therefore to try to contact creators directly and they may be more willing to allow their work to be archived as Dutch media heritage, but still it is difficult to persuade them what value this has for them. Unlike the relationship Sound and Vision has with broadcasters, this reaching out to YouTube creators is a very one-sided dynamic in which the archive has little to offer the creator other than preservation of their material.

In some cases Sound and Vision have established an ongoing relationship with creators whereby they acquire videos from specific popular Dutch channels on an ongoing basis. However due to the vast size and breadth of YouTube and the ambitions of the web video archive it does not seem that there is any easy solution to this problem of a one-sided and time-consuming process of acquisition other than to archive without permission. This is a wider problem associated with the archiving of social media and the web by centralised archival organisations when this activity can conflict with things like intellectual property rights.

YouTube

YouTube is a video-sharing platform and video repository. It was started in 2005 by three former PayPal employees who noticed a trend in people sharing video clips online (“YouTube - From Concept to Hypergrowth”). A secondary trend behind the success of the platform was the rise of smartphone and affordable video technology so that users could now create their own original video content and publish it online for a mass audience, leading to the tagline “broadcast yourself” being adopted.

As a result of these factors YouTube has grown in size and influence and now boasts over a billion users with 400 hours of video uploaded *each minute* (Brouwer). However, it is unlikely that YouTube even track this statistic (or care about it) on such a scalable and flexible platform. They prefer to focus on their user behaviour and the one billion hours watched daily in 91 different countries (“YouTube for press”). The platform also generates enormous income (as a result of these hours watched) for parent company Alphabet, as well as for many of its creators who take a share of advertising revenue and can earn a comfortable living if they build a significant audience following.

YouTube comments

YouTube’s comment section is a feature that has been included on the platform since its inception. It allows users to respond to videos, interact with video creators or with other users. A study of YouTube users also found that users read the comments for informational purposes but often are motivated to comment by a desire for social interaction, suggesting that there is no defined function for comments as such (Kahn).

Another key use for comments is by creators who can read, respond to, highlight and moderate comments beneath their videos. Comments allow creators to understand their audiences responses to their videos, interact with them directly and build a relationship and a following on the platform. This ability to interact and engage with audiences something many successful YouTubers have utilised in the comments feature. For example, PewDiePie is the most watched YouTuber with an enormous dedicated following that has grown over many years. It is clear from watching a few of his videos that PewDiePie reads comments and understands his audience. He also uses comments as a basis for new videos.

An example of using comments within videos is the format of reading and responding to comments, especially negative or “hate comments”. This format is common on the platform and an inventive way that creators deal with the problem of negative and offensive comments.



Fig 1. YouTuber PewDiePie is active in comments section. From a video titled “I actually read comments” (PewDiePie, 2018)

PewDiePie’s channel has also seen a controversy relating to comments and his community of followers. For example the casual use of racist language by commenters and by PewDiePie himself has led to criticism and an apology (PewDiePie, “My Response”). This type of thing is more familiar with the overwhelming impression many users have of comments is that they are low quality, offensive or spam.

YouTube has admitted that the comments sections are problematic. For example, an attempt to integrate Google+ (Google’s now retired social network) with YouTube was motivated by a desire to make comments more relevant so that users would see “comments from people you care about [rather than] whoever in the world was last to post” (Janakiram).

More recently YouTube was forced to disable comments on millions of videos across thousands of channels featuring children because some comment sections were being used by paedophiles on these videos. This shows the extent to which YouTube struggles to moderate comments at scale and instead is forced to completely disable them in order to ensure the safety of users . It is also worth noting that the real impetus for change with this example was advertisers pulling their ads from the channels. User comments on YouTube have long been problematic for user experience but now they are presenting a real point of financial risk. YouTube have something of a recordkeeping problem of their own relating to the appraisal of user comments.

Examples such as Google+ integration, moderation policies and processes, and the removal of comments also present a challenge for an approach to archiving this type of media.

Comments as archival records

There has not been any previous attempt to systematically archive YouTube comments as cultural heritage records. However, there have been projects setting out to archive social media, and other research and journalism projects which use YouTube comments as material which could be said to be a form of archiving (both covered in the next chapter).

An archival problem relating to comments is the concept of provenance which refers to the origin of a record or the context of creation, traditionally tracing it back to the individual, family or organisation that created or received the items (“Provenance”). This idea seems at odds with YouTube comments in a number of ways. For example the “original order” of comments is difficult to establish based on the point about YouTube being a fluid and changing platform made above. The origins of comments are also difficult to establish due to anonymity and a lack of a network or community that could indicate more about the context in which the comment was written. This is what YouTube referred to in their wish to make comments more like “conversations that matter” to users by making them related to a social network.

However, that does not mean that YouTube comments can’t be considered worthy of preservation. In fact, it may be the traditional notions of provenance that are limited. In the following chapter the case will be made that the significance of comments and their provenance can be established by understanding them as a response to the video. In this sense the video provides the provenance to the comments and vice versa. Interviewing people working on the web video collection at Sound and Vision the idea was suggested that comments could be considered as records of “audience reception” similar to the way in which Sound and Vision holds certain letters written to broadcasters (Kuypers). The following chapter will explore this idea as a possible framework for understanding comments, their provenance and their value as audience reception records within a media heritage archive.

Chapter 2: Comments as audience reception

This chapter presents a case for archiving comments at Sound and Vision as records of audience reception. This conceptual view of comments is a way to frame them as records of interaction between an audience and a media broadcast. It is also a continuation of the way in which Sound and Vision collects documents such as letters to broadcasters about specific programs.

Audience reception theory

Audience reception theory is a media and communication studies theory and methodology for analysing the way in which broadcast media is interpreted by consumers. Developed by Stuart Hall in the 70s it relates to the growing influence of television at that time and suggests that audiences play an active (rather than a passive) role when consuming mass media (Hall).

Hall identifies a disconnect between the production and reception of media communication. Using terminology from communication science, this is a distinction between *encoding*, meaning the production of a message, and *decoding*, which is the reception of the message by an audience. Hall suggested that the way in which a broadcast is decoded can vary. There are three ways in which a message could be decoded: dominant, negotiated or oppositional readings. The dominant reading is one which reflects the message intended by the producers. Negotiated reading is a middle ground whereby the viewer partly accepts the preferred reading. An oppositional reading is one that rejects the whole premise and offers an alternative viewpoint (Fig. 1).

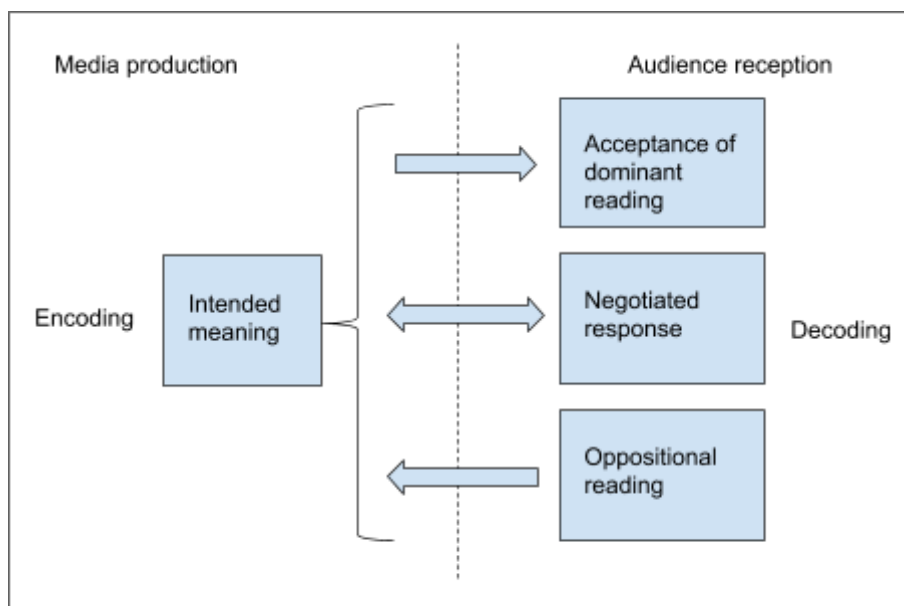


Fig. 1

An example of the application of this concept and method is a study conducted by Hall's colleague David Morley. A BBC news and current affairs programme Nationwide was shown to different audiences from different socioeconomic backgrounds. Morley attempted to understand whether the reception of the broadcast was in line with the dominant reading of the programme. The study found that certain groups accepted the dominant reading of the programme, while others only partly accepted (negotiated) it or completely rejected it (oppositional). The study demonstrated that a television broadcast doesn't necessarily have a fixed meaning and that it is hugely important to take into consideration the role that audience reception plays (Morley).

Literature review: Audience reception and social media

There are a growing number of examples whereby social media is collected and used to ascertain audience reception. Not all of the examples are directly influenced by the work of Stuart Hall but demonstrate the way in which these records can be said to relate to an interaction between audiences and media.

One example is a study of Tweets responding to Australian political panel discussion show Q&A. The researchers downloaded the #QandA hashtag using Twitter's API and then analysed the Tweets to see the extent to which users were engaging with the topic of debate. The hypothesis of this study was that information overload and the fast-paced nature of Twitter communication that would prevent users from engaging in "deliberative discourse" (Pond). However, the study found that the majority of Tweets did in fact meet their criteria for "deliberative discourse" whereby users showed a level of comprehension, empirical knowledge and judgement. This example shows a way in which social media records can be useful to determine media reception to some extent. The hypothesis of the study was that the nature of Twitter would impede the quality of discourse, but they found that this was not necessarily the case.

Another example of more systematic archiving of social media is the #Ferguson Tweet archive project and method developed by Ed Summers. Over 13 million Tweets from hashtags such as #Ferguson and related hashtags that followed the shooting of teenager Michael Brown in Ferguson, Missouri were archived. The method used the Twitter API and custom-built scripts for acquisition of Tweets at scale and then utilised the API again to provide access to the Tweets via a "rehydration" tool so as to avoid unauthorised publication of the dataset. This methodology is interesting in itself as an example of how acquisition of social media via an API comes with restrictions on how you provide access to that material and Summers developed an interesting workaround for that problem (Summers 2013). However, the process of "rehydration" also means that the archivist is dependent on the API and should Tweets be removed by users or by Twitter, then they are not retrievable (Summers 2014).



Fig 1. Tweet from #Ferguson that could be said to be oppositional response to what it suggests is the dominant media message. Images show people being treated after use of tear gas by police. High number of retweets suggests that it struck a chord with users

The #Ferguson Tweet archive can also be said to be media reception records and as having value as such. It has been noted that events like this one would previously have been ignored by mainstream media, but that social media platforms like Twitter mean that they now become big stories (Tufekci). The Tweets often question mainstream media coverage of the event and the aftermath (Fig 1.) and seek to give an alternative perspective to that offered by the traditional media. This in turn led to the hashtag #BlackLivesMatter and a wider discussion about institutional racism within the police and the media. This is the context behind seeking to archive these Tweets as important records of long term value.

Siersdorfer et al looked at YouTube comment ratings in order to discern whether a high variance of ratings can be linked to controversial videos and topics. Using a sample of 67,000 videos and 6 million comments they found that videos with high variance of comment ratings were also considered controversial by independent user evaluation. These polarising videos included clips relating to an Iraqi girl being stoned to death, President Obama and footage of the Tianamen Square “tank man” protest (Siersdorfer et al, 2010). They also found that high variance of ratings could be linked to polarising topics (which they identified using YouTube tags).

This example is interesting as it doesn't deal directly with the comments themselves but with the ratings of comments and suggests that this data can give some indication of audience reception. Users who view comments related to polarising videos that do not match their own view presumably seek to downvote those comments by way of negative ratings while other users will upvote if they feel the other way. This could be considered as an approach to using comments as audience reception to a video, but it is more complicated than that. Comment ratings are reactions to a user comment rather than the video itself so it is difficult to put them in the audience reception model as it has been defined here.

Another example - this time from researchers at Google - uses comments for the purposes of opinion mining (Severyn et al). This is basically sentiment analysis with regards to product reviews featured on YouTube. The problem the researchers focus on is the fact that "bag-of-words" style sentiment analysis fails to understand structural meanings of language. They apply a structural technique which differentiates between things like whether the sentiment related to the video creator or the product. This study also seeks to use comments as audience reception and acknowledges that there is complexity inherent in comments in that they are not all directly responding to the video itself, but may also be responding to some product or topic raised within the video. This highlights some potential limitations of basic sentiment analysis of comments that might directly relate the sentiment to the video.

Conclusion

These examples suggest that user-generated social media records like comments or Tweets can have some value but the relevance and significance of the records is really determined by the context.

For different research projects or decisions about archival appraisal the value and significance of comments depends on the context. For #Ferguson Tweets the Tweets serve as an important record that often contradict the mainstream media message or creates the message from stories that would normally be ignored. In the other examples comments and Tweets are used to understand how audiences are responding to a television program, what their opinions are about a product or whether a video is causing controversy. These are all different uses of social media to gauge audience reception within different contexts.

The question for Sound and Vision is what context comments would be of use to their organisation and users. I will be using the use case for comments as contextual records to go alongside their existing web video collection for the next chapter. However, I will also discuss this issue of appraisal and use cases in the final chapter of the report.

Chapter 3: Acquisition of comments via the YouTube API

This chapter looks at the capabilities of the YouTube API with regards to collecting comments for use alongside Sound and Vision’s web video collection. It will go over the basics of the API and some examples of it being used elsewhere, then explore the capabilities and limitations associated with the API. The chapter will then look at a case study in which the API has been used to archive annotations for 1.4 billion videos in 2019.

YouTube’s Data API

The YouTube Data API allows integration of YouTube features or data for the purposes of this project we will focus on accessing the data rather than using features via the API.

An example of using YouTube data with the API is the site SocialBlade.com which presents YouTube charts such as the “top 50 YouTubers” as well as dashboards (Fig. 1) for these channels all using data is taken from YouTube’s API but presented in a slightly different context. For the most part, this is for the purposes of comparing and ranking YouTube channels against one another and looking at channel performance over time.

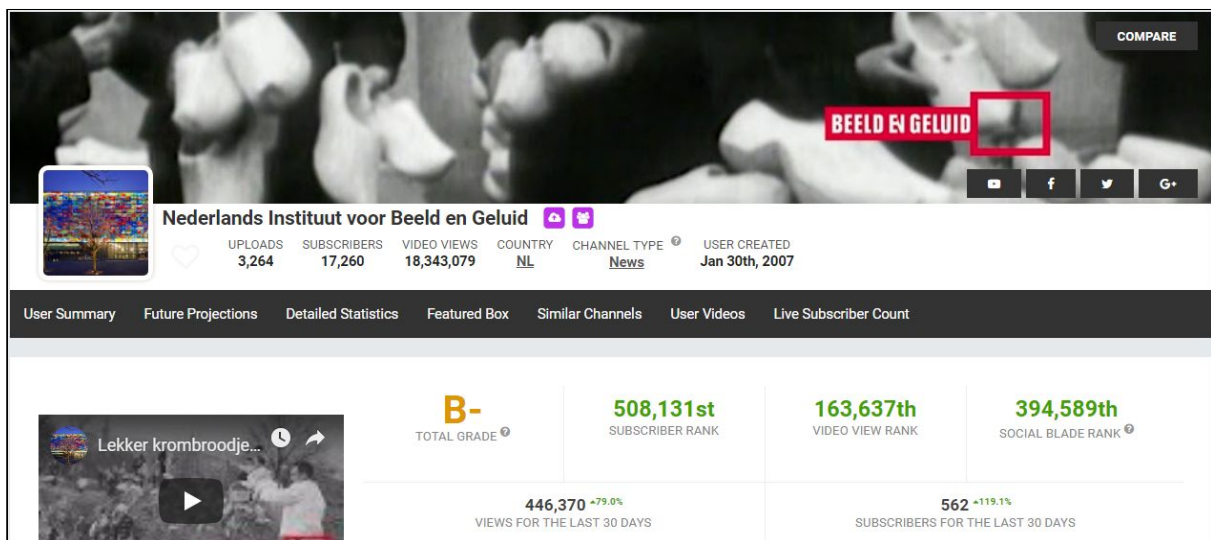


Fig 1. Beeld en Geluid YouTube channel dashboard on third-party website SocialBlade.com.

<https://socialblade.com/youtube/user/beeldengeluid>

Acquisition of comments via the YouTube API

Comments can be acquired via YouTube Data API v3 using a resource called “CommentThreads”. One way to access this without coding anything is to use Bernard Reider’s YouTube Data Tools at <https://tools.digitalmethods.net/netvizz/youtube/> (Reider). The VideoInfo function allows comments to be acquired by inputting a YouTube video ID.

CommentThreads takes top-level comments and replies for either a video or a channel. The number of comments underneath a video refers to both top-level comments and replies. For audience reception purposes it might be decided to only acquire top-level comments - not something that can be done using YouTube Data Tools without modifying the code.

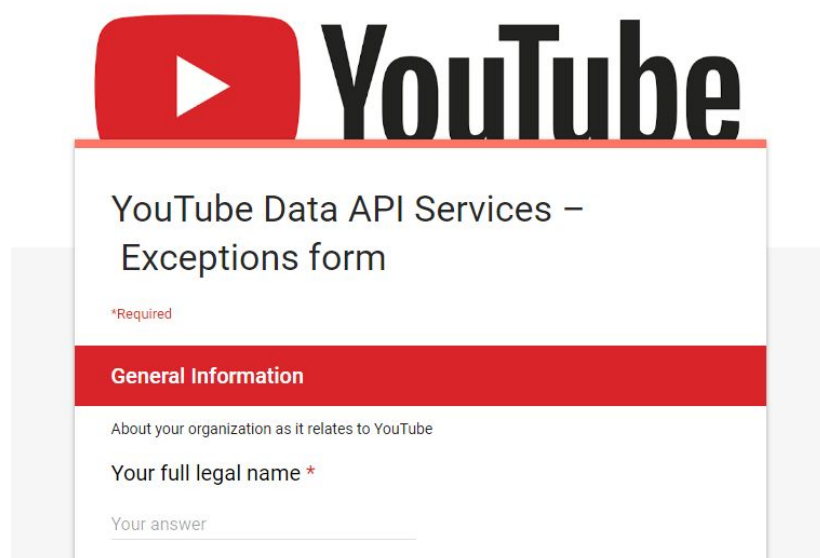
YouTube API quotas

Acquiring information from the YouTube API is done on a quota basis. The free daily quota for reading the API (as opposed to write operations which cost more) is 10,000 operations.

Requests for CommentThreads can use up to 4 units of your quota (2 for comment and 2 for replies). However, a request for replies may provide more than 4 comments. This is why YouTube Data Tools can get up to 100,000 comments because it utilises the CommentThread resource and the ability to get replies. So it is possible to get up to 100,000 comments within the quota, but many videos have far more than that.

It is also worth noting that the free quota limit used to be much higher at 50 million (“YouTube Developers Live: Understanding Quota in v3”). Since 2014 this has been reduced to 10,000. The reason for this reduction has not been discussed by YouTube Developers. It could be related to commercial interests relating to API usage but not necessarily. This highlights a risk associated with building a systematic process that relies on a third-party API which can be changed or cut off at any point by that party unless a formal agreement is made between Sound and Vision and YouTube.

The financial costs for increasing quota are not made available by YouTube. Users can apply to increase quotas at which point they are asked to set up a billing account. Upon setting up a billing account the user can then send a form to YouTube requesting an increase in quota. However, throughout this process there is no mention of pricing. The form asks for information about your organisation and how it relates to YouTube (Fig 2)



YouTube

YouTube Data API Services – Exceptions form

*Required

General Information

About your organization as it relates to YouTube

Your full legal name *

Your answer

Fig 2. Screenshot of form for requesting YouTube API quota increases

Calculating quota for full web video collection

One way of looking at collecting comments would be to collect comments for the existing collection of YouTube videos at Sound and Vision. This approach would mean collecting a large backlog that I estimate to be over 13 million comments in total¹. This could take between 130 to 300 consecutive days to download all comments, maybe more.

Another anticipated issue with regards to the quota limit is videos with very large numbers of comments. In order to do this, the comments would have to be acquired in stages, using the `nextPageToken` property as a way to start from where the last API request finished. Again, more information on this is in the API documentation, and it would be worth consulting a more experienced programmer and testing this out.

Acquiring 13 million comments would appear to be possible. It may not even require a huge level of time or effort. With some experienced developer resources the daily acquisition could also be automated in the form of a daily request to the API. However, this would need monitoring and tweaking and there are likely to be unforeseen steps that need taking.

¹ This estimate was made by pulling information for a large bulk of the videos in the collection from the API about the number of comments on each video and adding those up. The number of days is this number divided by the number of comments that can be acquired with 10,000 API units which seems to be between 30,000 and 100,000

Updating comments collection

Another issue is the fact that comments are continually changing. In some cases they are even removed entirely from videos by creators or by YouTube for safety or misuse reasons.

Decisions would need to be made about how this is handled and how this can be incorporated into the processes. It is possible that another script or program could be created to add new comments on a regular basis. This would then have to refer to the existing collection so as not to start requesting existing comments.

Distributed archiving with YouTube API case study



Fig 3. Screenshot from “The Time Machine” a choose-your-own-adventure YouTube series that uses annotations for functionality. Functioning version is available due to the work of the YouTube annotations archive:

<https://dev.invidio.us/watch?v=l8rJ1WML60Y>

Collecting 1.4 billion YouTube annotations

The YouTube Annotation Archive is a project that began on the /r/DataHoarder subreddit in December 2018 (“YouTube Annotation Archive”). This project began in response to the announcement by YouTube that it would be removing annotations from the platform within months. The community of r/DataHoarder were inspired by this and moved very quickly to begin collecting annotation data from the YouTube API using a systematic method distributed across a number of volunteer archivists.

YouTube annotations are boxes that appear as an overlay on a video. The ability to add them to new videos had already been disabled since 2017 because they are often annoying for users and also did not work on mobile. However there are many examples of inventive and interesting uses of this feature on YouTube and this functionality would be lost after the annotations were disabled. For

example, choose-your-own-adventure video series in which users can click annotations to decide different outcomes (Fig 3).

The annotation archive project ran over two months from the end of December 2018 to February 2019 and the final outcome of the project boasts annotation data and a separate metadata archive for 1.4 billion videos (“YouTube Annotation Archive - Final Update”). The bulk annotation files are hosted on the Internet Archive but are also made available via another API. This means the videos can also be accessed via YouTube mirror Invidio.us.

The methods of acquisition used for this project were complicated, involving specialised coding and scripts and a distributed method that avoided overloading bandwidth. This is beyond the scope of this report due to the expert knowledge required.

Luckily the project is very well documented. The best place to start is by looking at the Reddit posts (“YouTube Annotation Archive”) which also link to GitHubs and discuss methods. Reddit users related to the project on r/DataHoarders are also happy to discuss it, especially u/omarroth who was involved in both this archiving project and developing the Invidio.us YouTube player.

This example demonstrates how the YouTube API can be used at scale to archive a huge amount of data within a very limited time frame. 1.4 billion annotations were captured in about a month between the project being announced and the final post. This was done as a collaborative effort utilising the potential of Reddit and r/DataHoarder for sourcing volunteers to help out. Not only this but Reddit users contributed in the selection of videos that use annotation and so this also became something of a participatory project with fans and creators of annotated YouTube videos submitting requests for videos to archive.

Conclusions

Acquiring YouTube comments via the API is possible and with the right level of expertise the potential for acquiring large amounts of comments may be much higher than the 10,000 unit quota suggests. A specialised approach for collecting comments, even from videos with millions of comments, could be possible. If not then a script that can collect comments in stages and update existing comment collections seems completely feasible.

The annotation archive example shows what a group of highly motivated and well-organised “data hoarders” can achieve on a deadline. It may be worth getting a better understanding of the methods and legal status of this approach, as well as considering the collaborative and participatory nature of the project (many volunteers appear to have some stake as fans or producers of annotated Youtube videos).

Chapter 4: YouTube comments within FRBR-style Sound and Vision information model

This chapter considers how comments could be incorporated into the framework that determines the organisation of the Sound and Vision archive. This framework is based on Functional Requirements for Bibliographic Records (FRBR). Establishing how comments would fit into this framework is important as it determines how they will be organised within the archival system used at Sound and Vision if they were to be incorporated into the digital archive.

FRBR at Sound and Vision

Functional Requirements for Bibliographic Records is a framework developed by libraries and information services. It focuses on establishing relationships between bibliographic records based on the needs of users. Examples used to describe the framework often rely on records such as books within a library, but it is designed to work for a broad range of materials including audiovisual collections.

The stated goal is “to produce a framework that would provide a clear, precisely stated, and commonly shared understanding of what it is that the bibliographic record aims to provide information about, and what it is that we expect the record to achieve in terms of answering user needs” (*Functional Requirements for Bibliographic Records*). In other words, the functional requirements are determined by various user needs in relation to the bibliographic record.

FRBR outlines a hierarchical structure with four levels of description which correspond to four user tasks: find, identify, select, obtain. The levels of description are work, expression, manifestation and item (fig 1.). At Sound and Vision the user tasks and levels of description are slightly different to the original documentation.

The “work” is the abstracted level for an artistic or intellectual creation. The reason for this abstract level is that there is often no such thing as a definitive original work, only expressions of that work that may differ from one another in small or significant ways. From the existence of these expressions we conceive the abstract entity. For example, when we refer to Homer’s *Iliad* and Shakespeare’s *Hamlet* we are really talking about an abstraction rather than a specific expression.

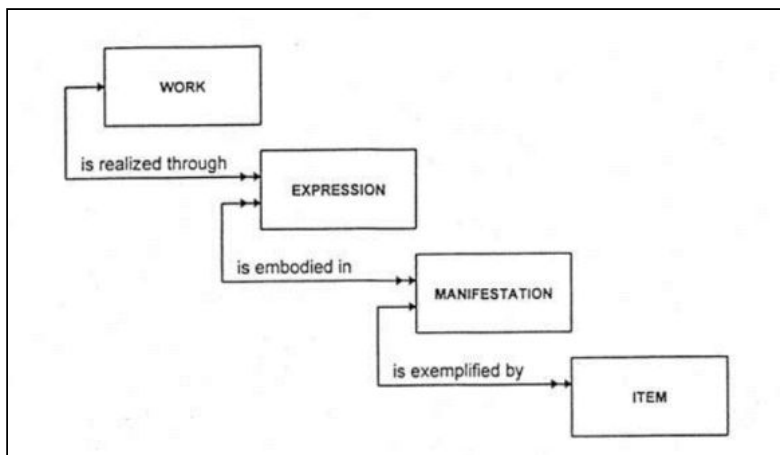


Fig 1. Group 1 entities and primary relationships (*Functional Requirements of Bibliographic Records*)

At Sound and Vision the “work” level can also incorporate a “series,” referring to the organisation of television programmes into series. The series is often the top-level of description for programmes within the archive. For audiovisual files that are not within a series such as web video then the series can be used more like a category with “web video” having its own categories such as “web video politics” (Fig 2.).

This is a departure from the original way that FRBR works whereby the “work” level describes an abstract entity and not a specific series or category. However, this is a broadcast archive and the contents and user tasks are different from the original use case that FRBR was based on.

Beneath the “work” level are “expressions” of the creative work. In traditional FRBR these can be different editions of Homer’s Iliad for instance, including translated versions. Expressions may also refer to a written play and a performance of that play as two different instances of expression.

At Sound and Vision “expressions” are the individual programmes themselves. This is the level at which the audiovisual files exist within the system. This is where Sound and Vision really deviates from FRBR, partly because this is a digital system in which it is not really possible to distinguish between the next three levels of “expression” “manifestation” and “item” as you can within a traditional bibliographic system.

Schematische weergave van het conceptuele iMMix-metadatamodel

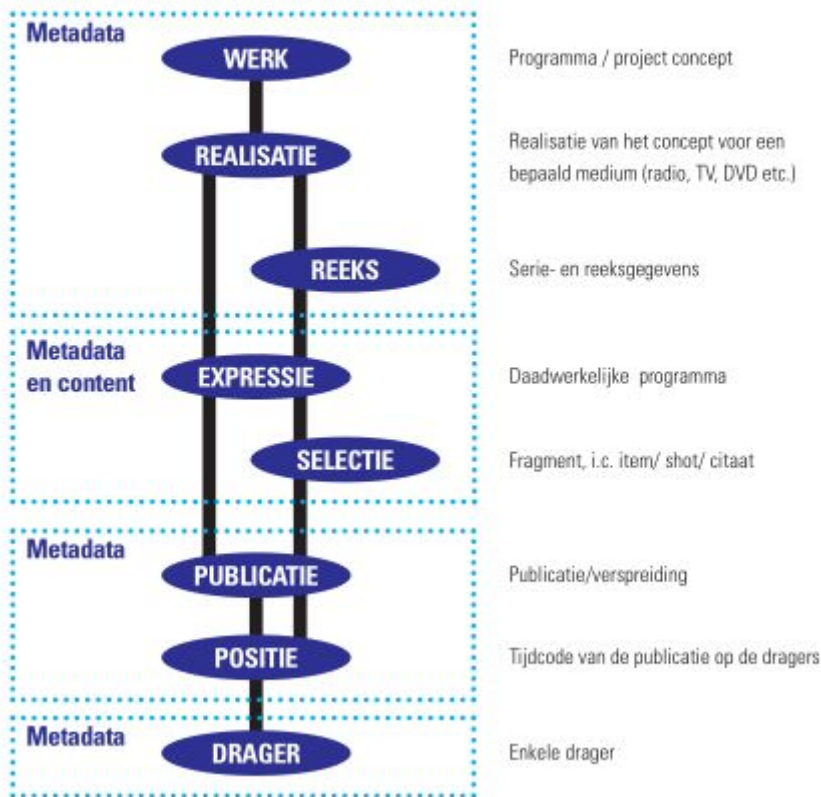


Fig 2. FRBR as used at Sound and Vision (de Jong, 2007)

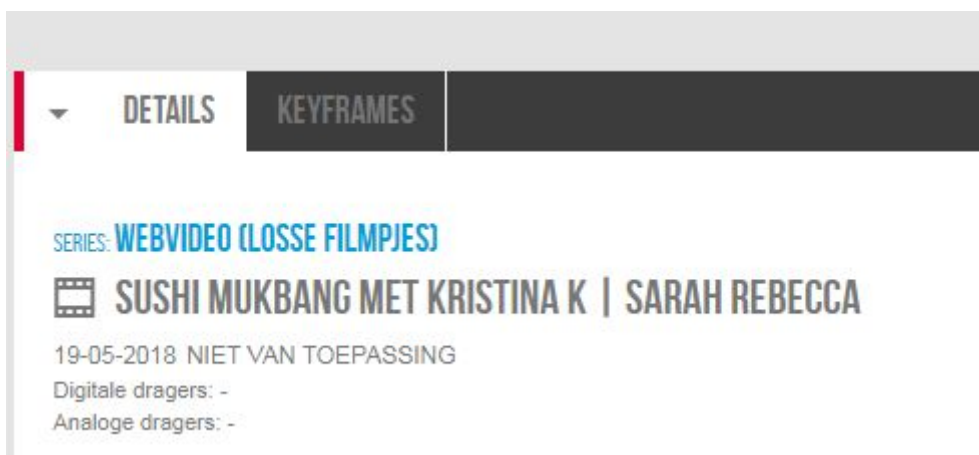


Fig 3. Screenshot from DAAN showing series “webvideo” for YouTube videos in the collection

In traditional FRBR the “manifestation” refers to a specific edition of a book or performance of a piece of music, but there may still be multiple copies of these which is why the next level is the “item” which refers to a concrete object.

At Sound and Vision both “manifestation” and “item” are treated differently, with the former being determined by the publication date at which the program was shown. In the case of YouTube this is different again and tells the date that the video was published. Sound and Vision also use metadata for timecodes within a programme, and this can be referred to at the “manifestation” level also should a media professional wish to access or refer to a specific shot or segment within a program.

At the “item” level Sound and Vision define the “drager” or “carrier” referring to the format. This makes sense for media professionals whereby the final decision is which format to obtain for re-use. For YouTube videos in the collection this is not possible because the audiovisual files have yet to be moved to the system (hence the missing information on the screenshot in Fig 3), but the planned format is MP4 for these files.

This use of the “item” level to determine formats is different to how the Library of Congress use FRBR in which the format is described at the “manifestation” level (Fig 4). This suits different requirements of the library which also provides access to physical entities such as books.

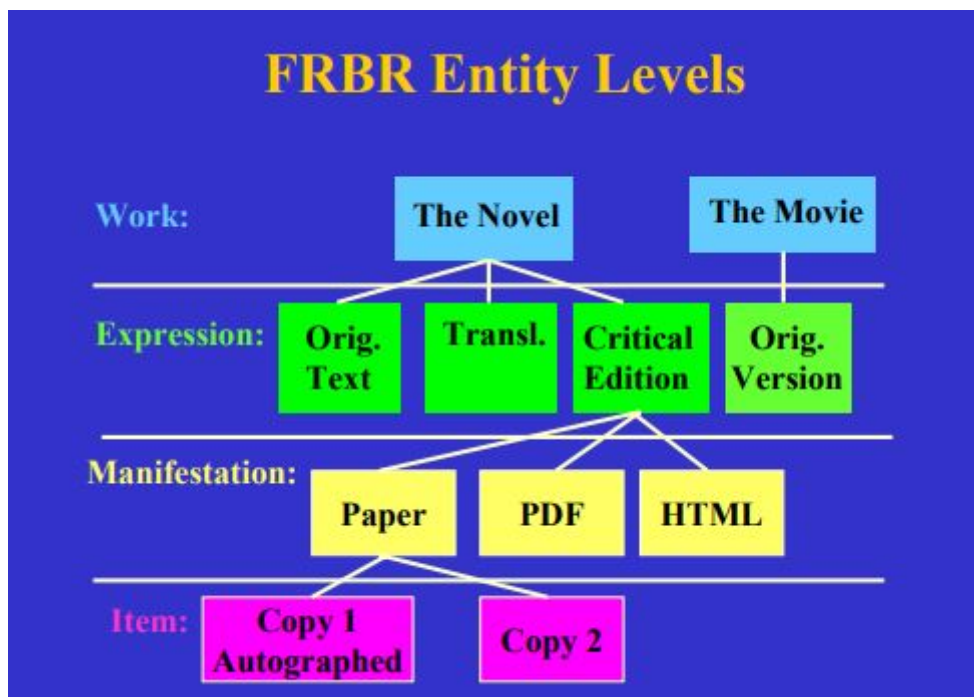


Fig 4. Library of Congress use of FRBR. Formats are established at “manifestation” level rather than the “item” level as used by Sound and Vision (Tillet)

This shows how the FRBR has been adjusted to suit requirements at Sound and Vision, primarily for the user needs of media professionals, who are the primary users of the system after all. However, it also shows some limitations with regards to web video, which is included as a category at the top level so that this level becomes a mixture of categories and series.

This brings us to the question of use cases for web video within the collection. Media professionals may understand how to search or even browse based on a series of programmes, channels, dates, or even by persons featured. But for web video the use case (which is what FRBR-type systems are designed around) is less clear (whether by media professionals or some other users) and that makes the organisation of that material less clear also.

YouTube comments

Having established the way in which FRBR is used at Sound and Vision, the next question is how YouTube comments would fit within this system. This presents a challenge due to the fact that audiovisual files are the foundation of the system. The slightly difficult organisation of web video also suggests that the system is designed specifically around broadcast media and even web video presents a slight challenge.

This also brings up the question of use cases again. The “functional requirements” that FRBR’s full title refers to are the requirements of the users. In the case of both web video and YouTube comments, the requirements of users have not yet been established and so it is difficult as a starting point. In this report suggestions for conceptual use case based on comments as records of audience reception have been made. The imagined user here is more likely to be an internal archivist working with a researcher to understand more about a YouTube video based on the comments at some future date when YouTube no longer exists. In this use case the comments must relate to the video within the FRBR system so as to make them accessible by referring to the same level as the video itself (the level of expression in Sound and Vision’s FRBR model).

YouTube comments as derivative and aggregate works

FRBR allows for relationships to exist between a “work” and derivatives of the work and based on the conceptual understanding of comments as audience reception they could be considered derivative works. This is a distinction from a model that would consider comments as a constituent part of the YouTube video.

FRBR also can be used to describe aggregate works. While the system was initially and primarily designed with the idea that distinct individual “works,” it has been shown that there is also a need for describing “aggregate works” (O’Neill). However, an agreed general rule for aggregates has not been

established, such as whether aggregation should take place at a “work” or “manifestation” level for example.

At Sound and Vision already uses aggregates in the form of a television “series” which aggregates different programs into a single “work”. Programs are then defined at the next level as “expressions,” as has been discussed already above.

Aggregation also offers a way to conceptualise YouTube comments in FRBR. Comments are not a constituent part of the video, so they would be considered “augmentations” as described in FRBR and part of a separate aggregate record that is derivative of the specific video to which comments relate.

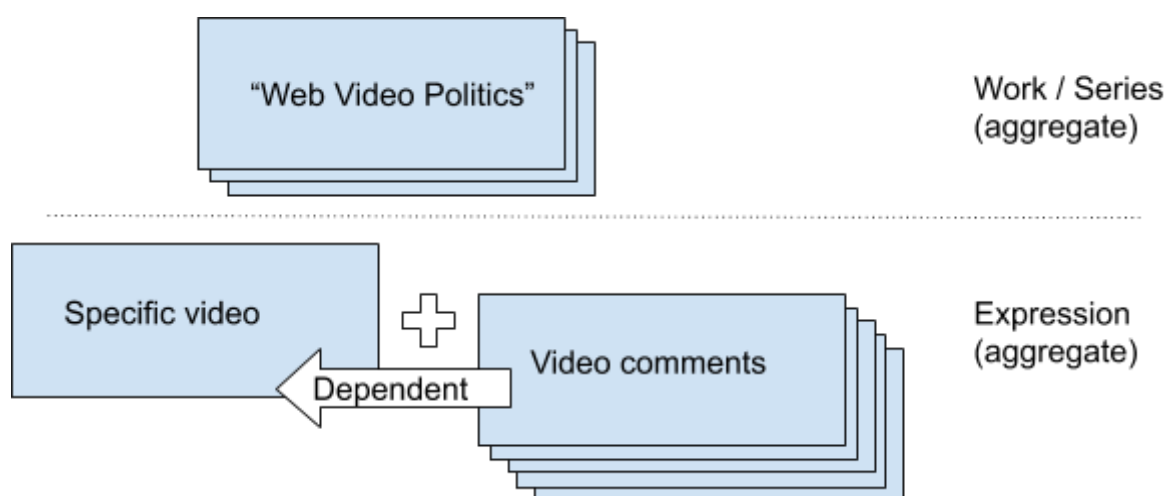


Fig 5. Diagram of how comments might work as aggregate records within FRBR-style system

Alternative approaches

The approach covered above is determined by the use of FRBR at Sound and Vision and the conceptual use case of comments as records of audience reception. There are limitations with this use case which is largely based on a hypothetical conceptual understanding of comments and an imaginary user seeking information about audience reception. With that in mind the following approaches are also potentially less complicated than including comments within FRBR.

Option 1: Collect comments and treat them as analogue for the time being

One approach is to consider how analogue items are handled by the institute. These also don't have a publication date, necessarily. For instance, Sound and Vision has letters within their paper collection that would relate to a certain episode or program. One example is the case of Phil Bloom, the first nude woman shown on Dutch television in 1967. This broadcast generated many letters of complaint which are preserved in the paper archives at Sound and Vision. However, currently there is no explicit connection between this material in the paper archive and the digital system. It is up to researchers to

find this material themselves. Perhaps this material could also provide useful context if it were highlighted or even digitised and made available via the digital system, but it is fairly likely that it is only of interest to very specific researchers who are happy to access the paper archives.

With this in mind YouTube comments could be treated like analogue collections. Kept and catalogued as part of a separate digital repository. If comments are associated with web video in the collections, users could be informed about this within the metadata or within the access tool for that collection.

The question then becomes about this separate digital repository. On the one hand, it seems like an unusual decision to make and could mean creating an entirely new set of processes and work, but if Sound and Vision wish to collect more non-AV material such as text and social media data in future then this offers much more flexibility, control and access options. However, this distances the social media data (such as comments) from the audiovisual repository. That being said, it is not difficult to imagine both DAAN and another system being brought together using their respective APIs. Comments and videos in both systems could be connected using the YouTube video ID which both elements share and is included in DAAN.

Option 2: CLARIAH Media Suite

One other system that already exists and is along the lines mentioned above is the CLARIAH Media Suite. This suite developed by Sound and Vision makes collections accessible to researchers, especially digital humanities and media studies researchers. It has its own API and different approaches to accessing the suite are being developed based on varying levels of technical knowledge so that researchers who can't write code to access the API can use alternative options such as Jupyter Notebooks.

The capability of the Media Suite with regards to YouTube comments has not been explored in this project but the platform is designed to offer features like "enrichment" of existing collections and it is being looked into by those more focused on research based use cases (Bernard Reider, forthcoming). An example of "enrichment" is the inclusion of ASR or automatic speech recognition as additional data. With this in mind it is possible to conceive YouTube comments being another possibility. It is not clear where the ASR data comes from and whether it is held by Sound and Vision or a third party.

The benefits of the Media Suite are that it potentially offers greater flexibility in terms of enriching existing collections for the needs of researchers. However the question of how ASR data is stored and preserved remains unclear.

Option 3: Host using an open third party

In the previous chapter of this report the YouTube annotations archive project organised by members of r/DataHoarder was discussed. This project is interesting as it demonstrates an approach to archiving that also provides completely open access to both data and methods (scripts used). Annotations were hosted by the Internet Archive but made accessible via an API. This API was then integrated into a range of third party YouTube applications such as Invidious so that annotations for billions of videos could be experienced by anyone. A clear issue with this is legality of making comments available due to copyright and YouTube's TOS. The open approach also has advantages such as the ability to provide access, integrate with third party via an API, and by increasing chances of developing interesting use cases.

Chapter 5: Discussion and legal considerations

This chapter will end the report with a discussion of the overall project and conclusions which will then lead into recommendations. Recommendations will be about next steps for this project as well as broader recommendations with regards to Sound and Vision’s approach to YouTube. This chapter is an opportunity to consider issues encountered with this project such as use cases.

Conclusions summary

The research question for this project was “how can YouTube comments used at Sound and Vision?”. The context behind the question is the transformation of the organisation from broadcast archive and museum to a wider scope that incorporates media beyond broadcasting such as web video and social media. This broadening of scope for Sound and Vision has led to experimental projects or “emergent practices” such as the video game archive project. YouTube comments present another opportunity for this type of emergent practice, seeking to test the waters with regards to collecting social media, but keeping within the scope of an audiovisual platform.

This project found that systematically collecting comments via the YouTube API is possible. It also found that they could theoretically be incorporated into the digital archive framework at Sound and Vision. The project also found that it may even be possible to provide access to comments based on the “rehydration” methods used by Twitter archives without breaching YouTube terms of service (see Ed Summers’ #Ferguson Tweets in Chapter 2). However, the project also struggled to define use cases and understand who the users of a YouTube comment archive would be.

Discussion: Issue of use cases

The decision was made early on in the project to consider use cases that were not media professionals or researchers. Media professionals and researchers are seen as having very specific needs and access to this type of collection did not seem appropriate to these users (Table 1).

Community	Example use case	Example of system used	Needs / goals	YouTube comments?
Media professionals	Re-use for broadcast production	DAAN	Access to high quality AV archives & detailed metadata	Seems unlikely they would be of use here
Researchers	Reference to	CLARIAH	Varied. Big data	Yes, of use, but

	material for research		analysis and digital humanities projects are mentioned	very specific needs
Education	Re-use of material for educational purposes	Education portals, school trips	Historical material, media literacy	Media literacy and media skills is a use case for comments (see below)
General public	Access to collections online	Websites, S&V YouTube Channel, Museum	Entertainment, media heritage, contextualised	Hard to see a use case
Internal users	Media management, metadata management, museum production	Various	Various	Context for collections as “audience reception” records

Table 1. Adapted summary of designated communities at Sound and Vision (taken from Verbruggen, 2015)

Possible use cases

From the literature research, interviews with staff and analysis of the API and comments data some tentative use cases can be suggested. However there is a big caveat around these which is that they are use cases without doing any real user research. As has been suggested, a weakness of the study with regards to use cases was to start with the material and then try to find a use case that would justify collecting.

Audience reception use case

One suggested use case already covered in Chapter 2 is the idea that comments are records of audience reception. As such they have a cultural heritage value and are an opportunity to store records of audience reception as well as broadcast output in the Sound and Vision media archive. The records could be used by researchers or internal archivists seeking to understand more about the context of creation for YouTube videos and the audiences that watched them.

Other uses of YouTube comments and social media for audience reception are discussed in Chapter 2 of this report such as the #Ferguson Tweet archive and the use of YouTube comments for sentiment analysis and detecting controversial topics. All of this could fall within the conceptual idea of comments as audience reception.

Historical context for a video

Another fairly researcher-specific use case is to understand the history of a video over time using comments. YouTube videos are not broadcast and watched all at once like many television programs. Instead they are watched over time by different audiences and the context they are watched in may change. For example the music video for Gangnam Style by the artist Psy is one of the most watched videos of all time. Many people watching it are more aware of this context than the original meaning of the video and references to the wealthy Gangnam District in Seoul that Psy is satirising in the song. This change in the audience's points of reference over time might be something you can see from the comments and may be of interest to some researchers. Unfortunately this was not something within the scope of this project. Gangnam Style has 4.9 million comments which goes well beyond the daily API quota made available by Google so downloading all of these would require a specialised script and methodology as discussed in Chapter 3 of the project. It was not possible to find a good example video with a manageable number of comments from which to build a potential use case for this idea in this project but could be interesting.

Community or participatory archive use case

When discussing YouTube it is easy to think of YouTube as the object of study or the thing that can be archived, but this perspective fails to appreciate that YouTube is not a singular entity. Unlike a broadcaster or even a network of broadcasters and production companies, YouTube is an enormous platform used by millions of creators. It's possible to think of YouTube as having certain characteristics, trends, and an internal culture, but this is a very limited perspective.

Another way to look at YouTube is to go beyond the platform level and to approach it at a channel level or community level. Many of the most successful channels work as a community in their own right or as part of a community. How comments are used and the type of users commenting varies between them.

Crowdsourced metadata

Another idea that came up in the project but was beyond the scope of testing out was the idea of using comments to generate tags and topic models for videos. YouTube's tagging and automatic topic generation isn't a strong point and YouTube are working on using tagging and topics more often because of the problems that completely automated solutions cause. There have been other studies

into this type of research using comments as a signal for determining types of video such as predicting controversial topics by comment behaviours as mentioned in Chapter 2 (Siersdorfer). The phrase crowdsourced here refers to indirectly crowdsourcing this from comments rather than making it a deliberate activity undertaken by users.

Alternative perspective: YouTube Creators are primary users of comments

This project was largely focused on looking at YouTube and comments from a distance or a top-down perspective and exploring possible ways to collect and archive these. Another perspective would have been to focus on the group that uses comments more than any other and for whom comments are an invaluable resource which is YouTube creators.

YouTube creators read comments to understand their audience and interact with them. Many of the most popular YouTubers such as vlogger PewDiePie build their audience and their popularity by being engaging and reactive to their audience. PewDiePie has grown a huge audience of 97 million subscribers over many years through consistently releasing videos and by building a community of engaged fans. He addresses his audience directly, refers to in-jokes, uses the audience to determine his videos such as the “Meme Review” series which is created from user submissions (admittedly via Reddit rather than YouTube’s comment section). He also creates videos in which he reads the comments and often refers to comments within his videos. From an outside perspective the importance of comments to PewDiePie and his channel is less obvious but it is probably fair to say that he reads comments extensively and that they are fairly integrated into his production process.

Another example of comment use on YouTube is the more recent story in which YouTube turned off comments across thousands of channels featuring children. This was done to protect users after stories emerged of paedophiles using YouTube and leaving predatory comments on videos of young people (cite). Advertisers also began pulling their ads from these channels and YouTube was forced to disable comments.

YouTube’s decision has been controversial with creators. Some have suggested the blanket ban of comments is discriminatory towards smaller channels. For instance, the channel “Special Books for Special Kids” has launched a petition to get comments reinstated on their channel. They argue that comments are key to the success of their channel and add a huge amount of value for their audience.

It is interesting that both PewDiePie and SBSK use Reddit as a secondary platform for audience interaction, showing the limitations of YouTube in this respect. The former uses Reddit for added functionality of submitting and voting on memes to be used for his “meme review” show. The subreddit has the advantage of being a more self-contained community of fans in this respect. The latter seem to use Reddit partly as a back-up community now that their comments have been shut off.

This shows that online communities extend beyond a single platform but also that different platforms can serve as spaces for commenting, audience reception and participation.

YouTube Creators are the new “media professionals”

This last use case focuses on the primary use of comments by creators rather than a re-use of archived comments. The reason for taking the discussion in this direction is that it seems to fit with the original purpose of Sound and Vision as supporting media professionals.

Sound and Vision primarily works with media professionals in Dutch broadcasting who use the collection for creating new programs. The shift of scope beyond broadcasting could also mean that the meaning of media professionals shifts also. In this broader frame a media professional can also be a YouTube creator. This media landscape in which YouTuber’s are new broadcasters is what Hilary Jenkins calls “participatory media” whereby technology has created a low barrier for anyone to become a media producer rather than just a consumer. The idea of the active consumer of media suggested by Stuart Hall has now gone a step further.

Chapter 6: Legal considerations

There are certain legal matters surrounding collection of YouTube comments which will be briefly discussed below. Unfortunately a more in depth look at legal considerations was beyond the scope of the project without consultation from legal professionals. However, it also is dependent on developing use cases. For some of the legal issues surrounding archiving user-generated material and online material is a grey area. Some organisations such as the Internet Archive take a very different approach than others for example..

Copyright and access

It may be possible - and my professor has argued - that YouTube comments do not fall within the scope of copyright law or intellectual property because comments are too short to constitute original works. For the majority of comments this may well be the case, but that doesn't mean that as a rule comments would not be copyrightable.

Publishing comments elsewhere would potentially be infringement of copyright on those comments. YouTube has a license to display the comments and even to make them available in bulk via their API, but that doesn't mean that any such rights are transferred to users of the API. In fact, YouTube's terms of service restrict the right to do so.

Presently Sound and Vision has a specialised license for acquiring videos from creators. Doing the same with comments on a large scale is not feasible and so any large scale collecting and publishing of comments data is potentially fraught.

However, the Ferguson Tweet archive presents an interesting approach by which users must "rehydrate" Tweets using the API (and their own API key) in order to access them again. Each Tweet has a specific ID and the accessible dataset constitutes a large downloadable dataset of these IDs. A similar process could be imagined for YouTube comments which also have IDs by which they can be called in the API. In theory access can be provided if those wishing to access get their own YouTube API key.

YouTube API

The archiving of YouTube comments using the API is heavily reliant on a third-party service that could be changed or terminated at any time. For systematic, ongoing acquisition which this project would seem to require, especially if comment sets are to be updated over time as well as continually

acquiring new comment sets for new videos, then there seems to be a potential risk or weakness in the process due to the lack of control over YouTube's API.

One solution is to not make this an ongoing systematic archiving approach and to make it a short term, one-off large-scale utilisation of the API which would probably require some sort of distributed archiving similar to the annotations project mentioned (or alternatively special allowances from YouTube).

Another solution is to consider YouTube as a partner and that the archiving of comments (or some other use of the API) can be mutually beneficial for both parties. This is why YouTube provides access to their API to begin with - to encourage experimentation and development of third party tools: "We provide the YouTube API Services to enable developers to create experiences that bring additional value to the YouTube ecosystem and its users" ("YouTube API").

This takes us back to the importance of use cases and the possibility of considering YouTube creators as media professionals. By serving YouTube users and creators, and understanding YouTube's organisational goals and the problems they have (some of which have been outlined in this report), there may be a possibility to partner with them.

Working with an archival institution could also help YouTube's public image and be valuable to them at a time when they are under fire for spreading misinformation and failing to keep control of their system. Recently notable changes have been made to remove offensive channels and changing algorithms so as to reduce visibility of misinformation on the platform. In a 2018 interview CEO of YouTube Susan Wojcicki said she thought of it as a "library" and wanted to promote the value of YouTube as a platform for learning and information (Thompson) demonstrating a desire re-brand as a more trustworthy information service. Maybe YouTube needs to start thinking like an archive as well.

Chapter 7: Recommendations

This chapter sets out some recommendations based on the research project. These are fairly loose recommendations and more like suggestions for next steps and ideas

Recommendations for API and acquiring comments

- Programmers needed. Systematic archiving of comments at scale via YouTube's API requires making scripts that can deal with quotas and collect comments in stages
- Contact YouTube via the form mentioned in Chapter 4. They may be able to provide a higher API quota for this project
- Be aware that relying on API means that access is not always guaranteed. A possible solution is to develop a relationship with YouTube based on a use case that adds value to their users and creators

Use cases / users

- Start with users and user research. This project looked at technical and archival requirements but did not do user research or seek to understand users (e.g. YouTube creators)
- Idea of YouTube creators as new "media professionals" and potential users of archive

Archival methods

- Macroappraisal - Functional analysis. The second chapter on audience reception can be seen as a starting point for thinking about appraisal of records based on the functions of those records. This method called functional analysis was not known at the time of doing the project.
- Participatory archiving - Another approach to archiving online communities is participatory archives whereby the community and users are actively involved in the archiving (e.g. appraisal, arrangement) and will make decisions that work from a record creator / user perspective. This could be applied to YouTube communities. YouTube's ability to organise material is very limited and often people are better at organising this type of content than machines (see McCulloch article about *Archive of Our Own* use of tags)

Works Cited

- Brouwer, Bree. "YouTube Now Gets Over 400 Hours Of Content Uploaded Every Minute." *TubeFilter*. July 2015. <https://www.tubefilter.com/2015/07/26/youtube-400-hours-content-every-minute/>
- de Jong, Annemieke. "Metadata Model Beeld En Geluid Biedt Gebruiker Van Digitale Content Diepte En Structuur". *Informatie Professional* 06 (2007) <https://publications.beeldengeluid.nl/pub/454>
- Functional Requirements for Bibliographic Records*. International Federation of Library Associations and Institutions. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- Hagedorn, Berber. "The End of the Television Archive as We Know It? The National Archive as an Agent of Historical Knowledge in the Convergence Era." *Media and Communication*. Volume 4, Issue 3, 2016. Pages 162-175
- Hall, Stuart, and Morley, David. *Essential Essays*. Durham: Duke University Press, 2019. Print.
- Janakiram, Nundu. "Turning comments into conversations that matter to you." *Official YouTube Blog*. November 2013. <https://youtube.googleblog.com/2013/11/youtube-new-comments.html>
- Khan, M. Laeeq. "Social media engagement: What motivates user participation and consumption on YouTube?" *Computers in Human Behavior*. Volume 66, January 2017, Pages 236-247
- Kovach, Steve. "Alphabet had more than \$70 billion in market cap wiped out, and it says YouTube is one of the problems". *CNBC*. April 2019. <https://www.cnbc.com/2019/04/30/youtube-algorithm-changes-negatively-impact-google-ad-revenue.html>
- Kuypers, Valentine. Personal interview. 16 May 2019.
- McCulloch, Gretchen. "Fans Are Better Than Tech At Organizing Information Online". *Wired*. June 2019. <https://www.wired.com/story/archive-of-our-own-fans-better-than-tech-organizing-information/>
- Middag, Ineke. Personal interview. 16 May 2019.
- Morley, David. *The Nationwide Television Studies*. London, Routledge, 1999. Print.
- O'Neill, Edward. "FRBR Aggregates: Their Types and Frequency in Library Collections." *Association for Library Collections and Technical Services*. Vol 59, No 3. 2015.

Pond, Philip. "Twitter time: A temporal analysis of Tweet streams during televised political debate."

Television & New Media 17.2 (2016): 142-158.

Rieder, Bernhard (2015). YouTube Data Tools (Version 1.10) [Software]. Available from

<https://tools.digitalmethods.net/netvizz/youtube/>

Severyn et al, "Opinion Mining on YouTube". *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. June 2014. 1252-1261.

<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/42471.pdf>

Siersdorfer, Stefan. "How useful are your comments?: analyzing and predicting youtube comments and comment ratings." *Proceedings of the 19th international conference on World Wide Web*. April 2010.

Summers, Ed. "A Ferguson Twitter Archive." *inkdroid*. August 2013.

<https://inkdroid.org/2014/08/30/a-ferguson-twitter-archive/>

Summers, Ed. "On Forgetting." *Medium*. November 2014.

<https://medium.com/on-archivy/on-forgetting-e01a2b95272>

Thompson, Nicholas. "Susan Wojcicki on YouTube's fight against misinformation." *Wired*. March

2018. <https://www.wired.com/story/susan-wojcicki-on-youtubes-fight-against-misinformation/>

Tillet, Barbara. "The FRBR Model." ALCTS Institute on Metadata and AACR2. April 2003.

<https://www.loc.gov/catdir/cpsol/frbrengr.pdf>

Tokmetzis, Dimitri. "Radicalization on YouTube: we investigated hundreds of thousands of videos and millions of comments". *De Correspondent*.

<https://decorrespondent.nl/9151/radicalisering-op-youtube-zo-onderzochten-wij-honderdduizenden-videos-en-miljoenen-comments/445626247-173eb676>

Van der Vliet, Martijn. Personal interview. 22 May 2019.

"API." *Lexico*. <https://www.lexico.com/en/definition/api>

"Comments", *YouTube Data API*, <https://developers.google.com/youtube/v3/docs/comments>

"CommentThreads," *YouTube Data API*,

<https://developers.google.com/youtube/v3/docs/commentThreads>

"My Response." Uploaded by PewDiePie. 12 Sep 2017.

<https://www.youtube.com/watch?v=cLdxuaxaQwc>

“Provenance.” *Society of American Archivists*.

<https://www2.archivists.org/glossary/terms/p/provenance>

“YouTube Annotation Archive - Final Update.” *Reddit (r/DataHoarder)*. March 2019.

https://www.reddit.com/r/DataHoarder/comments/b7imx9/youtube_annotation_archive_annotation_data_from/

“YouTube Annotation Archive”, *Reddit*,

https://www.reddit.com/r/DataHoarder/comments/aa6czg/youtube_annotation_archive/

“YouTube Annotation Archive”. *Reddit (r/DataHoarder)*. December 2018.

https://www.reddit.com/r/DataHoarder/comments/aa6czg/youtube_annotation_archive/

“YouTube Developers Live: Understanding Quota in v3.” *YouTube*, uploaded by Google Developers, 12 February 2014, <https://www.youtube.com/watch?v=JNEJPffn1k>

“YouTube for press.” *YouTube*. <https://www.youtube.com/yt/about/press/>

“YouTube - From Concept to Hypergrowth - Jawed Karim”. (Re) Uploaded by Lucio Amaral (originally uploaded by Karim himself but since removed). 31 Jan 2016.

<https://www.youtube.com/watch?v=7oJdD2oUHXc>