

# Scaffolding Community Documentary Film Making using Commissioning Templates

**Tom Bartindale**

Open Lab, Newcastle University  
Newcastle upon Tyne, UK  
tom@bartindale.com

**Guy Schofield**

Theatre Film and Television,  
University of York  
York, UK  
guy.schofield@york.ac.uk

**Peter Wright**

Open Lab, Newcastle University  
Newcastle upon Tyne, UK  
p.c.wright@ncl.ac.uk

## ABSTRACT

Crowdsourced video is now a viable tool with which broadcasters and communities alike can produce authentic, high quality video content. However, the literacy, language, skills and tools to produce a documentary through commissioning content are currently difficult to acquire. We explore opening up the documentary film commissioning process to community contributors by developing a framework which instructs, guides and informs non-professional contributors in capturing the content required for making videos. Through the results of an in-the-wild deployment we discuss how our framework scaffolds content creation, the capture of high quality footage and coordination amongst teams of contributors. We then discuss how this can inform community media creation in the future.

## Author Keywords

Coordination; video; community; film; cinematography; media;

## ACM Classification Keywords

H.5.3 Group and Organization Interfaces – Collaborative Computing

## INTRODUCTION

Research has shown that mobile and web technologies have provided myriad opportunities for the general public to quickly and easily create video content [5]. Mobile devices equipped with cameras are becoming increasingly ubiquitous and have potential to not only capture high-quality footage but also facilitate sharing and communication between users. User-generated video produced by non-professionals is rapidly emerging as a key part of the professional media production landscape. This is driven in part by broadcasters wanting to harness the power of the crowd to deliver a richer documentation of events through an audience's perspective. It is also driven by the availability of community

dissemination platforms which empower communities to document events in a way that professionally commissioned content could not, for example, by being able to capture content from a large number of different locations at different times [19]. In addition, new emerging models of production such as those identified by Green *et al.* [11] call for re-thinking the traditional workflow, and confounding traditional constraints in order to support new modes of media production and consumption.

There are, however, two key obstacles to be addressed in order to fully exploit the potential of community driven video, both for amateur and professional practices. The first is that community produced content is often not of a high enough aesthetic standard to be used in professional quality documentary film. One response to this problem was the Bootlegger platform [16], which aimed to help novice contributors compose shots that could be used professionally, using only mobile phone cameras. As reported in [14] while the platform was successful, it was limited to supporting the production of music videos. A second and related obstacle is that in the crowdsourcing model, professional producers have no means of directing potential contributors in exactly what needs to be captured. They are therefore forced to sort through and edit large volumes of video to find useable material post-capture. A potential solution to both of these obstacles would be a system that allowed commissioners to set out the high-level narrative for their video project by selecting collections of shots from a library of genre-specific options. In this model, each shot would be accompanied by guidance and support to enable contributors to both capture the required subjects but also to meet professional aesthetic standards.

In this paper we propose a framework of 'commissioning templates' which embody the crucial information (drawn from professional practice) required for basic film production. The framework scaffolds the creative, logistical and technical requirements for producing documentary video which can be of value to both communities and broadcasters. Templates can be easily designed by non-professionals, and help community contributors to democratize both the commissioning and capture process, whilst retaining the core values of shot quality and content which are key to producing high-quality documentary content.

An online template designer supports novice commissioner's work, beginning from the defined genre conventions of documentary film, and the resulting template helps novice contributors capture footage which matches these conventions. Through implementing and deploying a web and mobile platform built around generic customizable commissioning templates, we aimed to develop an understanding of templates as a method of scaffolding community created content.

In developing the system, we envisaged non-professional communities coming together either independently of -or in collaboration with- professional commissioners, to define what content would be useful and interesting for them. They would do this in a way that could be communicated simply to shot contributors and support them in delivering both higher quality and more viable footage. We discuss the design of commissioning templates and their implications for community content commissioning using analysis of data collected during an 8-day deployment at a large international arts festival in which an untrained volunteer 'camera crew' captured a range of arts events. We initiated this shoot alongside a major broadcasting organization as an exploration for how community teams could commission and contribute to a wide variety of documentaries about the experience of visiting the event.

We worked with our commercial partners to understand the challenges they face in creating content at large-scale events and we discussed the value of commissioning templates for representing key parameters of film making, and how these could work to communicate often complex requirements for footage. In understanding these existing production and commissioning constraints, our aim was not to replicate existing workflows, but to explore alternative workflows by which communities can themselves produce and manage content which is of a standard that is acceptable to our the professional broadcast companies.

To conclude, we argue that there is value in creating a space in which novice communities and professional broadcasters can communicate with a shared understanding of these requirements. This opens up new media paradigms in which broadcasters can commission documentaries directly from communities or communities can shoot documentaries which are subsequently disseminated by broadcasters. Finally, we raise key questions about the nature of film literacy, emergent behavior and the importance of audio in community film making.

## PREVIOUS WORK

Festivals are an interesting test-bed for our system because they are an example of complex large-scale events for which broadcasters typically have difficulty documenting the participant experience, but where community captured content has significantly increased. As summarized by Bulterman *et al.* [5], the landscape of authoring community based content has dramatically evolved over recent years, with a trend of increasing literacy in using camera phones for

video capture and streaming [7] alongside improved quality in mobile capture technology.

Previous work has presented solutions for compiling and browsing content [19] crowdsourced from the community in these scenarios however much of this work has centered on post-capture media authoring rather than supporting the commissioning and coordinating of UGC content capture process to generate viable original footage. Various attempts have been suggested to bridge the gap between professional and community production to produce high-quality documentary content, specifically through scaffolding (providing a structure which communicates or parameterizes the tasks and skills needed to produce content that is useable) sections of the production workflow. Automatic algorithms [16], peer-to-peer awareness [15] and community contributed meta-data [10], have all been proposed as methods of guiding capture and sense-making from communities, however little has been proposed to scaffold contribution to community commissioned content and how to support the commissioning process itself.

## Community Sourced Content

A number of commercial products demonstrate the validity of leveraging crowdsourced media post-event and support novice users in compiling media. The Fan Footage [21] platform and similar projects [18] allow contributors to submit their own videos of a concert for aggregation by the band, and in return their footage can be played alongside a high quality audio recording, and the Burst [22] platform supports contributors in real-time streaming of situated events so organizers can leverage multiple viewpoints for broadcast. Although these platforms use efficient methods of aggregating and selecting content, the decision on what to capture is made by the contributor without guidance or coordination, resulting in content which lacks continuity and varies in quality and upon which it is therefore difficult to extract or impose any structured narrative. Guimarães' *et al.* investigation [13] into how the relationship between contributors and the subject of capture affect editorial decisions suggests that a variety of factors (such as intimacy, the amount of effort required to capture and privacy) impact on the decisions made in what is valuable content.

## Commissioning Content

Platforms such as Seen It [23] have begun to explore delivering a more structured approach to commissioning, allowing a producer to ask for specific sequences which can be delivered by community contributors. This top-down pre-planning approach integrates well with professional broadcast workflows and consequently has been used to commission audience point-of-view content for broadcast [24]. However, the task of communicating shot requirements necessitates that the producer provides instructions which are then interpreted by the person capturing. Additionally, this simplified pre-scripting approach does not take into account teams of contributors working towards the same show, and

is unable to distribute shot allocations depending on location or teach non-skilled contributors what is required.

Real-time co-ordination as a method of organizing the capture of content has emerged in response to such team based scenarios. The Bootlegger [16] platform aims to capture specific sets of shots by replicating the role of a live camera director, coordinating teams in real-time by showing contributors on-screen overlays for shot framings and making sure specific genre conventions are maintained. In comparison, CoStream [15] uses peer-to-peer awareness to allow contributors to make decisions on what to shoot based on the knowledge of what is already being recorded, and SwarmCam [8] supports sports fans in delivering live coverage of ice-hockey games using mobile devices. In comparison, rather than coordinating the definition and capture of footage using pre-scripting or real-time communication, ‘Run Spot Run’ [10] uses community input to tag runners in marathon footage both in real-time and in post-production, and ‘Watch-and-comment’ [6] and ‘Rock N’Roll’ [17] supports community tagging of videos to help highlight interesting segments and build a community understanding of valuable content. Each of these systems demonstrates how we can leverage existing media literacy and mobile technology to support community contributors in producing quality content, but all fall short in providing ways for communities to commission their own content for capture.

### Defining Commissioning Requirements

When collaborating on film projects, industry professionals use a common vocabulary to communicate ideas quickly and concisely. This vocabulary is used by producers and directors to [2] specify and commission content and coordinate film crews, communicating key elements such as mood, style, aesthetic, subject and genre by describing specific parameters of film making such as (shot type, length, focus, etc. To communicate these requirements in industry, mainstream broadcasters have a specific set of criteria [4] for determining if content is viable for use. Although the criteria for user-contributed content is more relaxed, style, quality and framing are key measures of clip viability. Engström *et al.* [9] clearly demonstrates how teams of novice participants can coordinate themselves to produce live event footage, but that a tension exists when they consider themselves as audience members as well as camera operators and as such, framing decisions are made differently. Although it has been argued that new forms of literacy skills will emerge for these types of mobile video recording [20], the current genre

conventions of documentary film demand that footage is shot in particular styles, lengths and framings.

### COMMUNITY SOURCING CONTENT

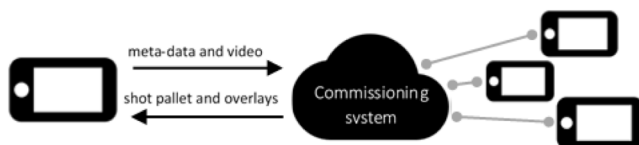
Looking to the professional and broadcast model of video production, the workflow is clearly compartmentalized, with specific areas of skill and responsibility led by highly trained people. This professional workflow is resource intensive, limiting the breadth of content that a professional organization can capture. At community level, this often means local news or documentary topics are overlooked.

A common example of this problem is the filming of local marathons by national broadcasters. Understandably they cannot spare significant resource for complete coverage of an event that lasts several hours and takes place over a wide area, given that their aim is to produce a 1 hour live show. This often results in coverage being limited to celebrity or professional runners. Additionally, the editorial constraints and nature of the end-result specify limited parameters for capturing additional content (such as interviews with local people, or a spectator’s viewpoint). Recently broadcasters have begun to explore using content from amateur contributors, briefed before the event (as with Scene It), or aggregated post-event (as with Fan Footage, wall of moments, and projects sourcing from social media). Classically this strategy is framed as ‘crowdsourcing’ or ‘user-contributed’ content [12], and in most cases involves giving little guidance (if any at all) in specifying what video should be captured and submitted. As discussed, the crowdsourced model assumes no feedback or understanding between the *contributor* and the *commissioner*.

We propose that a more appropriate model is ‘community sourced’ video. In a community-sourced scenario, an individual or group would act as *commissioner*, describing the collection of footage required for the video as a whole (not just individual clips) to produce their required output (video edit). This idea of content being delivered in response to ‘self-directed commissioning’ is in contrast to the traditional ‘broadcaster led commissioning’ model. The individual or group would act as a combination of director and producer by setting the tone of the footage, and designing, planning and specification of shots for a shoot. These detailed requirements would be put out for contribution with the aim of capturing enough relevant and contiguous content to produce a more complex and compelling final video than conventional crowd sourcing. Contributors would capture content against the requirements, but without (necessarily) having knowledge of the final edit or the state of production. The resulting footage would then be aggregated and shared with the commissioner as raw files with meta-data describing each contribution against the request, and traditional editing and post-production workflows would be used to complete the project.

### A COMMISSIONING TEMPLATE

We worked with our broadcast partner to understand the constraints on commercially viable content and the variety,



**Figure 1** Bootlegger platform supports requesting content from contributors’ mobiles and aggregating resulting clips

quality and style of content that could be successfully integrated. In a professional scenario, commissioners' requirements and creative decisions would be communicated and provided on location by a producer (working with a trained camera person). However, in a community scenario, a digital representation of the footage requirements is needed. Text descriptions (used in Seen It), and visual overlays (used in Bootlegger) have proved successful in defining requests frame by frame. However, these do not specify the precise subject, quantity of shots and context the footage needed, and Mascelli's [25] '5 C's of Cinematography' (camera angles, continuity, cutting, close-ups and composition) present a method of parameterizing such information. The template therefore needs to efficiently represent the requirements for the final edit, the tone and style of the footage and the decisions made by the producer/director. To represent this complex set of criteria, we define each commission (or *shoot*) as a *template* with associated *parameters*. When filming, contributors select from a list of *roles*, which are arbitrary groups of *shots* pre-prepared by the commissioner. These can be used to represent geographical locations (e.g. outside the venue, at the after party), punctuation in events (e.g. before-gig, during-gig) or other semantic information (e.g. interviewer role, on-stage role). Each individual *shot* is a definition of what a contributor can capture. Each *shot* consists of a *frame overlay* image, a title, *description*, *maximum length* and quantity *wanted* (figure 3). The *wanted* parameter is designed as a guide for a contributor to keep track of how many *shots* of each type they are capturing to guide them in exploring alternative angles. Taken together, these parameters allow the commissioner to define footage shot-by-shot, whilst seeking to maintain continuity, tone and quality.

### SYSTEM FOR COLLECTING MEDIA

Building upon the open-source Bootlegger [26] mobile video platform, which provided an infrastructure for distributing individual shot requests and aggregating captured media from participants. This tool was designed for live band recording, and lacked tools for users to easily create and manipulate commissioning templates so we implemented a prototype online template editor. Contributors log in via a modified version of the Bootlegger mobile application which presents them with a list of shoots to contribute to Obtained from the server. After selecting their desired *role*, the contributor is presented with an on-screen list of requested shots. Selection of one reveals the visual overlay, text description and record button. When recording a *clip*, a progress dial displays the record length / max-length, and cuts off the recording when the maximum is reached.

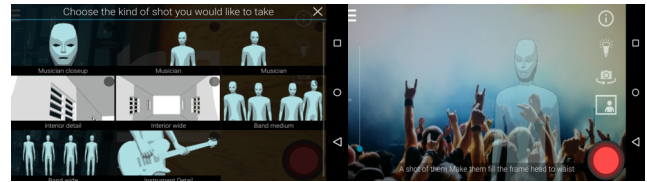
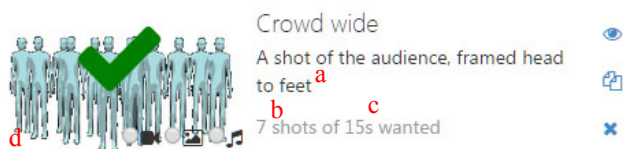


Figure 2 Left: Shot selection display, Right: Frame overlay and description

Figure 3 Close-up of editing an individual shot definition, a) description, b) wanted, c) maximum length, d) overlay frame

Upon recording a clip, the mobile device uploads a static thumbnail and meta-data (containing which *shot* the clip was taken of, who, where and when it was captured). The user can later choose to upload the original content on a suitable data connection.

### DEPLOYMENT SCENARIO

Through our broadcast collaborators, an opportunity was presented to document an event from an attendee point of view, by deploying our commissioning and contribution system at the Edinburgh Fringe Festival, billed as the world's largest international arts festival. The festival encompasses myriad forms of art and entertainment, crucially from a wide cross-section between amateur and professional acts. Because of constraints on resources, the our partners can only record a tiny subsection of the festival (<0.01% of over 3000 shows) and only from a 'professional' perspective, and so our drive was to explore the feasibility of creating a canonical video encompassing smaller acts, street performers and the wider festival visitor experience. This would depend on obtaining footage that could not have been reasonably captured using the broadcaster's traditional workflow and without experienced crew, financial or logistical support. We would obtain footage through: i) shooting shows from the point of view of audience members; ii) capturing filmed interviews with performers across the city; iii) capturing street performers and buskers; iv) capturing footage of the city, venues and general festival experience, all of sufficient quality to be potentially used by our professional collaborators. Amateur volunteers would be required to both *commission* and shoot the video (*contribute*), or to *commission* for another team to shoot. We deployed an amateur film crew of 15 participants recruited from our institution (students and staff), each spending 2-3 days over an 8 day period on location at the festival, participating in exchange for free access to shows. Most of the participants had limited experience in using the mobile application and online template designer, only 2 had professional film experience. 30% of participants had produced or shared video within the previous 6 months. Prior to the deployment, we contacted a variety of venues and acts, preparing a calendar of ~2 shows and ~2 interviews per day. The brief given to each participant on arrival was simple: make a film about your experience of the festival. No further information on cinematography or editing practice was given. Each day started with a team debriefing of the

previous day's activity and logistics session for the coming day (recorded), after which participants were split into teams of 2-4 depending on agreed schedule of events. Each team was then given ~2 hours to design templates for shows that day, alongside monitoring and maintaining a 'Buskers and B-Roll' template used by the whole group for street based shooting, and reviewing their footage from previous shows. 2-3 hours were then spent filming street performers and interviews, before teams went to shoot their shows. Roles and shots were chosen by the *commissioner* (team) responsible for the shoot, but because teams were re-using and adjusting templates throughout, these roles rarely changed.

### The Shooting Workflow

The online template editor provides a pre-developed list of simple documentary templates designed by the authors based on well understood genre conventions, which acted as a starting point for the film [2]. Users could then adjust parameters, edit or remove existing shots, or add additional ones from our archive of ~150 overlay frames (figure 4), sourced from well-known cinematographic texts to create a template which represented their commission. Using this workflow, we envisaged specific alternative combinations of use: i) a team designing a template for themselves, ii) a team designing a template for another team, iii) an extended, or iterative process of maintaining a template.

### INVESTIGATION METHOD

Our participants attended 14 live shows, conducted 16 interviews with performers and recorded 2 live music events across traditional theatres, coffee-shops and bars. Additionally, 140 min of street performers, street interviews, cutaways (shots giving context to a narrative e.g. a close-up of someone's hands handing out a flyer) and b-roll (contextual footage, e.g. people queuing for a show) were collected. We collected the template designs and clip meta-data for each shoot alongside a range of quantitative data and logs of system usage. The resulting data consisted of 23 completed template designs against which 15 participants captured 4869 clips (of which 1685 were not uploaded due bugs in the prototype). In addition, we performed a guided

group interview each day, resulting in 1h45m of transcribed discussion. Various statistical analyses were performed on the data in relation to specific parameters in the template framework. The resulting metrics form the outline categories for our findings. These were then interpreted using the results of a thematic analysis performed on the qualitative data, and grouped into three areas of interest: shot template parameters, template construction and usage and collaborative insights. Throughout, we were particularly interested in the decision process usually performed by the camera crew and director/producer, exploring the relationship between commissioning and contribution through the lens of template design, and participants' experience of these processes.

### FINDINGS

On average, 4 people contributed to each (specific event) shoot, whilst 15 people contributed to the collaborative template. The mean length of a clip was 22.84s ( $\bar{x} = 15.41s$ ) and the mean number of clips contributed per shoot was 235. The quality of footage produced is fundamental to determining its value, particularly when understanding footage shot against a given template. To ground any conclusions drawn from the analysis, we randomly sampled 5 of the 18 shoots. For each shoot, we assessed the quality of the footage according to standard editing criteria (in focus, not shaky, good picture quality, enough length to use), guided by the broadcasters' technical standards (Figure 5),

Shoot	Total Clips	Total Uploaded	Useable of Uploaded
1	17	17	15 (88%)
2	19	10	10 (100%)
3	23	13	13 (100%)
4	167	164	157 (96%)
5	292	222	146 (66%)

Figure 5 Report of viable content in a sample of shoots

with the understanding that the majority of footage matches the requested overlay-instruction [16].

This high-level analysis demonstrates that approximately 90% of footage captured is viable for use, arguably more than a conventional production model might capture for a final edit. A key factor to consider given these results however is the varied quality of audio obtained from mobile devices. In many cases, ambient factors prevent audio from being viable for editing, although it can be used for continuity and identification purposes. Given this inherent constraint of mobile hardware, clips were not rejected solely on their audio quality. The relatively high quality standard of the contributed content informs our understanding of both the validity of the template structure and the participants' experience of shooting and designing templates.

### Interpretation of Individual Shot Descriptions

As discussed, the key element of a template is the *shot* which consists of a *frame overlay* describing the required framing,

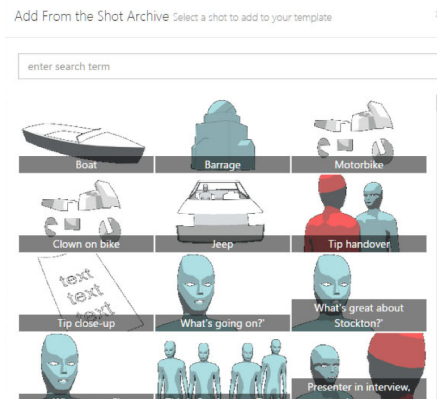


Figure 4 Archive frame selector from which commissioners can select additional framings.



a *description*, *maximum length* and *wanted* number of recordings.

#### Which shot overlays were used and how?

Although Bootlegger offered 120 framing overlays, only 20 were included in templates by commissioners. Figure 6 presents the distribution of *frame overlays* chosen across contributors for those 20 overlays. The area of the each circle represents the proportion of shots that were of a particular overlay. For example, contributor 1 used relatively few overlays, and used overlays 2,3,14 and 15 (respectively: figure-mid-shot, figure-medium-shot, stage-mid-shot, figure-close-up) most frequently. Also as can be seen in Fig 8, overlay 1 (guitar-wide-shot) was used by very few contributors, except for contributor 6.

Looking across the graph, what is evident, is that although each contributor had preferred framings, most contributors took a range of frames. This data clearly shows overlays 3, 9 and 13 (which map directly to the top 3 shots recorded overall) were being favored by most contributors, but the reasonably consistent variance suggests the majority of contributors were choosing to shoot a variety of different framings, making use at some point of most of the available options.

Looking down the graph, clearly some shots are favored and some are hard to take. This can in part be attributed to the limited scenarios presented but is more than might have been expected from novice contributors, particularly considering a traditional documentary comprises ~6 common shots [2]. Indeed, one aim of providing a palette of shots from which contributors could select was to encourage capture of a range of framings, rather than just focus on those that are easy to capture or easy to understand.

The two most common framings were medium-shot: 38% and low-medium-shot: 19%. This is understandable as these framings are easy to obtain on a mobile device, while maintaining a ‘socially acceptable’ distance from the subject, or whilst sitting down in the audience. Promisingly however, the 3<sup>rd</sup> most used shot overlay is a close-up (11%), a more difficult shot to obtain on a mobile device and arguably a less

natural shot for a non-trained contributor to capture, suggesting that the overlays are encouraging contributors to capture shots outside of their normal practice. This is supported by one contributor’s interview comments: “*I started to think, "That would make a good shot". I was trying to then match it...I felt like I was only starting to think like that during the process of using the template...*”

A number of more subtle factors emerge from the qualitative data influencing which framings contributors chose to make use of. In many cases, users were making abstractions and assumptions about what the shot they were being asked to capture, sometimes interpreting the shot in terms of the wider context of what the shoot is trying to capture: “*...I used the medium landscape shot because you want the crowd. It's about festivals and this is somebody amongst a crowd.*” And in some cases abstracting the shot overlay visually, in terms of optical characteristics of the requested frame, such as a contributor referring to an overlay of close-up hands: “*I was kind of interpreting it as a close up of that sort of area, or just them doing something.*” In most of these cases of abstraction, the shot meaning was maintained, but in some, the framing was so abstracted that the associated meta-data lacked meaning: “*We did Venue Exterior which did give us a lot of freedom of what we were filming because it's a landscape. We weren't really following a template with that.*” Although sometimes struggling, contributors learnt that a variety of shots were possible from limited seating positions. Interestingly, one contributor commented that they were treating the overlay as the most important part of the shot request, and was aware that often what they shot did not correspond to the context description which accompanied it: “*[It's important for the editor to know], "Oh yes, this a medium shot," and looking for the medium shot. It's because maybe I used overlays, but without filming the category of the overlay.*” The particular configuration of venues often limited their choice of shot, but through the demonstration of possible shot framings, they developed an understanding of different possible options, in fact suggesting that providing a set of possible options increased their awareness of shots to take and encouraged them to try new ones.

#### What lengths of recordings were they shooting?

A key parameter of a shot definition was the *maximum length* of a shot allowed. This was introduced to limit the resulting footage to significant valuable segments under the assumption that contributors would record for longer than requested if not limited, which would result in an editor navigating large amounts of contiguous data in post-production. The graphs in figure 7 show the lengths of clips captured for those shot descriptions defined with 4 different *maximum length* parameters. What these graphs show is the degree to which contributors chose to terminate the recording before the system terminated it automatically at the given *maximum length*. For 25s, only 18% of clips were terminated before the upper limit, for 60s 48%, for 90s 89% and for 120s, 86%. Of particular interest, for 90s and 120s, most clips were terminated ~25s suggesting 25s was

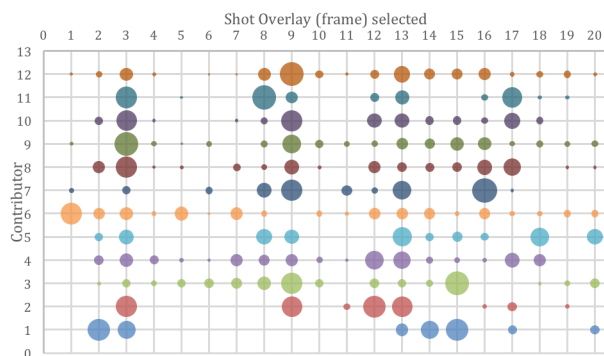
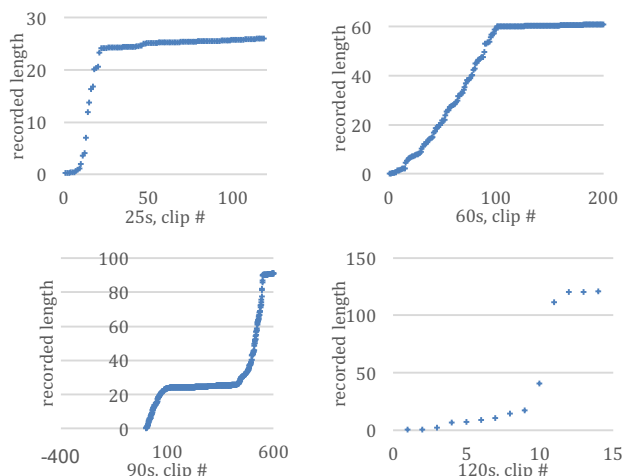


Figure 6 Number of frame overlays captured by each contributor

considered as an optimum length of clip, but that there was value in allowing contributors to record longer clips. The original imposition of the limit was to avoid capturing redundant footage, but as we see from this graph, when given the opportunity to record longer, contributors did not reach the limit in most cases. The fact that such a low limit was hindering contributors is reinforced by the qualitative data, and we can see that a number of factors impacted on this pattern of shot lengths. A problem with the default shot length (15s) was identified by contributors early on. They felt that having such short clips led to a very repetitive experience where they were frequently selecting a shot and pressing record to capture content. This was identified as an issue particularly for interviews which required some continuity of recording: “[the length of shots is] actually a big deal when *Bootlegger* stops or you're not fast enough with the next shot.” In response to this, template commissioners increased the default **maximum length** to 90 seconds or higher. The resulting templates were felt to give more creative control to the contributor, allowing them to stop the recording when they felt the subject had been



**Figure 7 Actual lengths of captured clips against maximum length of shot**

captured properly. This is supported by the resulting length distribution for longer **maximum lengths**.

*What tensions arose for contributors as audience members?* An important factor in both deciding what shots to select and how long to record for was the nature of contributors being audience members and camera crew. One reported that the palette of shots was: “... quick enough to navigate, but sometimes I'd miss a good joke or something because I'm too busy clicking” Whilst acknowledging that their role was to capture an audience perspective, they felt creatively constrained by having to maintain their seats. When opportunities arose to move around venues, they were taken and used to deliver ‘better’ footage from an audience perspective: “I was ... not surrounded by any other audience members. So I just felt like a lot freer, than I have done so far, to actually move and things. So I felt like I could capture

*better as an audience member, what I wanted to capture, without having a bit of a paranoia.”* In order to understand what appropriate footage to capture however, it was important for contributors to be aware of the content: “Yes, you need to engage slightly in order to understand what would be good.”

Contributors also reported that there was a clear tension between discerning what content to capture based on your creative judgment, and making sure to capture enough for an editor to make the decision later: “You can't be an audience member either. The whole point is that you're showing other people something you like or you're a fan of.” This tension between audience member and camera crew raises a number of questions about the nature of performance capture by audiences, but also how to legitimately balance understanding what to record of a context with concentrating on the act of recording. Indeed, some contributors felt under pressure to capture as much as possible for continuity, particularly before the shot lengths were increased: “I was pressing the same shot, mainly because I couldn't get any other angles because of the type of environment.” Having experienced this tension first hand, some commissioners started to be concerned with how contributors would be able to deliver on particular shots given the shoot context, and how the act of shooting would impact on the rest of the audience; “[it] doesn't feel right that someone's kneeling down right in front of the stage trying to get nice close ups without a zoom.” and considering the impact on the rest of the audience in what shots would be requested: “...people are going to be sitting on round tables so you don't want to be standing on chairs getting high shots like that, or right in people faces.”

### Template Construction and Usage

Having reported on the impact and interpretation of individual shots, we can explore how the design and use of combinations of these shots within templates impacted the workflow. Templates were designed by a variety of contributors for both themselves and each other. Overall, 18 templates were designed, and as mentioned, each resulted in a mean average of 235 clips per shoot captured against 19 shots by 4 contributors.

#### *How were templates constructed?*

As discussed, templates were constructed by selecting a relevant starting template from the archive provided and building upon it. From this early point, commissioners were making use of their tacit and experiential knowledge of similar types of events to predict the venue conditions, and often started by searching in internet for photos of the venue. “Yes, I think given that we would probably be seated all around the venue what sort of shots would be possible from there?” However this was done without any real thought to narrative structure of the resulting footage: “We didn't have a plan like this is our storyboard. We picked random ones and thought we'll probably use this one.” Structure did emerge though in terms of logistical planning in an effort to

simplify their own planning, such as separating groups of shots for interviews into different roles than those for a show. Once a routine of creating templates for shoots was established, templates from previous shoots (from any member of the crew) were cloned and adjusted: *"It was like a [previous] one and then we modified it. A lot of them were musicians or close up on guitars, so we removed those ones. [we added] general close ups, performance pieces."* In many cases this adjustment was only related to specific shots relating to unique elements of that show, and the main body of the template remained. *"I just took out the instruments and stuff when we were filming plays or whatever. I just went to the archive and tried to find general shots that might make sense, like close-up interior."* Acting as both commissioners and contributors, they transferred shooting experience to make adjustments to future template designs, learning what shots were practically possible: *"We chose to put that in yesterday and it worked. We said it was like a black box, so we had that stage-wide black box thing"* and creatively possible: *"[I remember] shots that I thought would be relevant today which I'd never really thought about before from a film making perspective."* After a few iterations of this "clone and adjust" strategy, a standard template emerged which contained their collective interpretation of what shots were possible in most venues and feasible for contributors to capture, however shot decisions were not only influenced by the practical considerations, but also by producing footage of a particular aesthetic appropriate to the subject matter: *"[I thought] you don't necessarily want loads of extreme close ups of hands and stuff like that... Because there's not that much to see in terms of that."*

#### *How did potential editing influence template design and contribution?*

Alongside logistical, practical and experiential factors influencing template design, commissioners started to envision how footage might be used in an edit and the editor's experience. The question of how much to capture was raised by contributors, considering the possibilities provided for editing in both style and the role of continuity. For example, in describing how they had improved upon the shared 'Buskers and B-Roll' template, two contributors chose to add specific shots with an intention of how they might be used in an edit: *"We expanded and are trying to get more of the landmarks and details interspersed between shots."* As well as appreciating the editor's role in interpreting the footage, there emerged a tacit understanding of what might be interesting to a potential consumer, whose when describing their experience of watching live show recordings raised the question: *"Why is the camera just focusing on this screen, it looks terrible"* However, a tension remained between capturing everything vs. capturing interesting bits and commissioners had different approaches, such as capturing as much as possible: *"I felt like I wanted to get the whole thing. If I was reporting and showing other people what was interesting, I felt that I needed the whole thing."* Or capturing only what is useful for an editor: *"But*

*are you doing a favor by filming the whole thing or would it be best to film just six or seven and be content with that because when you have too much footage then it becomes [too much to edit]"* In some cases, they would capture what they thought was interesting and limit footage for the editor, in some they would capture as much as possible to let the editor decide, and in other cases they would either communicate with each other for continuity, or look to see what other people were capturing to calculate whether a shot was already covered or whether it was important or could be ignored.

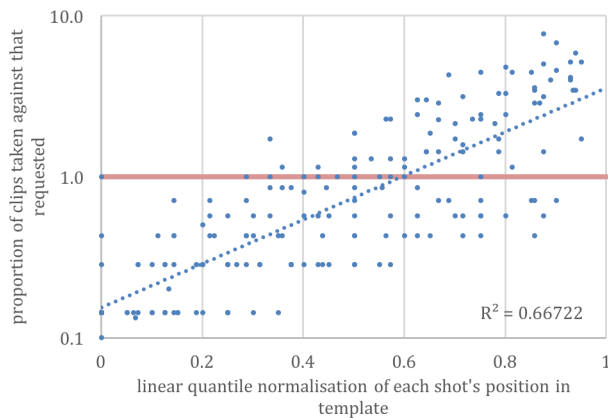
#### *How does the number of shot definitions in a template affect capture?*

We have already identified that not all shots were captured by all contributors, and that many were over-subscribed. Commissioners were free to include any number of shots in a template, and in fact a mean of 34% of shot definitions were not used at all for a given template. When the proportion of unused shots in a template was plotted against the total shots available in the template we expected to see a proportional relationship, indeed one contributor considered templates with fewer shots to be clearer in what they were requesting, and easier to understand: *"Yes, I think maybe even a bit clearer than yesterday because there are less options."* however as  $r^2 = 0.1614$ , this suggests there is no correlation between how many shots were provided and those actually used. A proportion of each template were ignored and this could be explained by the shot selection UI of the mobile application, which displays 9 shots being the user has to scroll, and that contributors were unlikely to scroll and explore the available options (possibly due to time constraints) before selecting a shot. Indeed, one contributor recalls that they *"...missed the first bit because I was trying to find the right shot."* Additionally, feedback suggests that in some cases, shots that were deemed to be similar by the contributor were treated as one: *"Well I had another shot like that so it kept asking me to do that and I just thought it didn't really exist, like it wasn't relevant."* or should be ignored as irrelevant due relative to the contributors location or context.

#### *What was the impact of the wanted parameter?*

Most documentary film includes a variety of shot types but some types require more footage than others (e.g. shots of a presenter). The **wanted** parameter was designed to allow commissioners to specify to contributors how many of each shot type were desired. On the contributors' application, this was represented by a changing visual indication of how many clips of each type they had already captured. We know that the feasibility to match a shot frame was one factor in contributors choosing what shot to select. We would expect that contributors would show some response to the **wanted** parameter through what shot decisions they made. However, when comparing the quantity of shots captured relative to the quantity **wanted**, as in Figure , a clear pattern of over-subscription emerges. This data shows ~50% of shot definitions in a template were over-subscribed across the whole template (where the intersection  $y=1$  equates to clips





**Figure 8 Proportion of shots captured relative to the wanted parameter**

taken=clips wanted and lie along the red line). Given the logarithmic scale (and  $R^2$  value of 0.66), this shows a significant over-capturing of video in around half the requested shots. Interestingly, for any given shot, a mean average of  $2.7\times$  ( $\tilde{x} = 0.57\times$ ) the *wanted* parameter were actually captured. A number of possible reasons for this disparity could be considered given the qualitative data; the *wanted* parameter could not have been clearly communicated to the contributor through the mobile UI, or the wanted parameter was not set to a value that made sense for the contributor in the scenario, as exemplified by one contributor who felt duplicates were occurring: “I got to the stage where I felt like I was repeating similar things I’d already done.” Alternatively, contributors could have used their own judgement to override the parameter, capturing what they considered to be shots of justifiable quality and framing. In fact the decision on what to shoot regardless of the *wanted* variable was in some cases influenced by the perceived use of the clips for an editor: “I ended up just not even trying to record some of it, just because it seemed a bit pointless.” This either limited their recording of shots that had been requested, or in some cases shot definitions were ignored, just so that content that was deemed important was captured.

#### *Does pre-planning foster contributor creativity?*

Learning from experience, commissioners tried to manage how much and what contributors were asked and reduce the on-site decision making that needed to take place, particularly to streamline time sensitive situations: “...there wasn’t much time and [what we found was], the best thing to do is storyboarding. ...You have a few drawings that tell you, “This is the angle. This is the angle. This is the angle” and the job is done.”

Both shots and positions where contributors could stand were pre-planned in this way. This pre-planning has the potential for limiting the creativity of contributors, however presenting contributors with a pallet of shots (even a limited one) chosen by a commissioner fostered the perception of increased creativity for one contributor who recalled, “...just feeling like you can do what you want a bit more, and you’re

shaping what you’re getting a bit.” Indeed, defining groups of shots which are seen to be important to the commissioner still leaves the contributor to make the final creative decision on what to shoot: “Yes I think the structure “Do this, or do that shot” [makes it less stressful and] In a way it lets you be a bit more flexible a bit more and you can be more creative.” In fact, even when the template had been designed with generic shots which didn’t apply, contributors still managed to capture creative shots, as they used the list of shots as a reminder of the variety that was expected: “Still having the prompt...I did use at some point when I thought it was relevant. So it’s a good reminder of the variety of shots that you could take.” Limiting the shots in a template could however be conceived as limiting contributors in what they can shoot, resulting in them capturing less footage, or producing content with misrepresentative meta-data. One contributor showed annoyance that shots were not available in the template when they were actually possible, emphasizing that context of shots is important in designing a template: “Yes but also if I’m at a gig where I can get to the front... I want to be offered then, whereas if I’m sitting in an auditorium and I can’t stand up, I don’t want to be offered then.” However, when reflecting on the how the template performs in more spontaneous situations, one contributor felt that reducing the number of shots available to shoot limited their responsiveness to spontaneous events.

#### **Collaboration and Crew Usage**

A key value of the template system was to communicate commissioners’ requirements to multiple contributors, who were sometimes co-located.

#### *How are contributions distributed across individuals for a shoot?*

The template is designed to share shot allocations equally between contributors. As mentioned earlier, on average 4 people contributed to a shoot. One interesting question is to consider if people contributed equally to shoots. The average variance in the proportion of shots taken by each contributor in a shoot is 0.117, indicating that for any given shoot, users on average contributed to within ~10% of the mean contribution proportion. This low variance in proportional contributions suggests that in our case at least, there was no noticeable over-contribution by particular people for a given template, and that work is distributed evenly.

#### *Can contributors trust a commissioner’s decisions?*

In some cases the template may be the only communication between commissioners and contributors about a shoot. How contributors interpreted the template therefore was based partly on their trust of the commissioner. In most cases, template decisions were not questioned, but as contributors became more experienced, they perceived a lack of control over what was in the template, making it restrictive: “I think, for me, it would feel better as someone taking the shots, if I had that control over them. Rather than someone having semi-autonomy as a director.” In most cases however, having the shot decisions made by someone else took the pressure off the contributor in choosing shots, “When we did

*the interview shot we didn't have to go through all of the shots and find it. It just make it a bit clearer and easier."*

#### **What is the value of continuity?**

Through a combination of reviewing both their own and other's footage after it was shot, and the experience of contributing, the crew realized that continuity was a significant issue for the team, which was not covered by the template structure. Two strategies emerged for maintaining continuity in scenarios where it was deemed important by contributors: i) making sure to keep recording as much as possible: "*[I had to press record] to make sure there was as little gap between the recordings as possible.*" and, ii) developing simple hand signals to coordinate contributors to film continuously: "*Yes, we had a very simple system to kind of alert the other person...so we were trying to make sure that we were both shooting at the same time.*" Interestingly, these hand signals show similarities to the signals developed by professional film crew at live events [14].

#### **DISCUSSION**

Our aim was to investigate the feasibility of scaffolding community-sourced documentary through the use of parametrized templates that describe the footage required. In our investigation we asked key questions about the structure of templates and how they were appropriated by contributors. Specifically we explored how the overlay, maximum length and wanted parameters worked together to enable the description and capture of specific shots. We also enquired about how commissioners constructed templates and how they aided the coordination of teams of contributors to deliver appropriate content. Commissioning templates are inherently a prescriptive way of defining and describing requirements for video footage. However, given that mainstream documentary makes use of a relatively narrow set of shots and angles, providing pre-seeded templates was found to be helpful to commissioners in planning relatively complex shoots. In fact, through alternating the roles of contributor and commissioner, our volunteer users translated their practical experiences of shooting into improving templates. Specifically the process of contributing to templates informed them of the logistical and practical limitations of specific contexts and how to plan accordingly. Rather than limiting the creativity of contributors, the use of a variety of shots in a template encouraged contributors to attempt new framings. Interestingly, individual contributors did not feel limited or restricted by the available options, rather they abstracted meaning from framing overlays, and once the collective decision was reached to increase shot lengths, they used their own judgement and creative awareness in shooting appropriate lengths and quantities of footage without these limits being imposed. The subtle uses of templates that took place suggest that the contributor's inherent media literacy was being drawn out by their use. The perceived reduction of risk for the contributor in making creative decisions about what to shoot, as demonstrated by the disparate shot counts relative to the wanted parameters suggests that they were able to use their personal judgement

about what was valuable to capture for the editor. The quality and practicality of audio capture on mobile devices was raised a key point of concern by both contributors and commissioners. Increased media and technology literacy has clearly impacted on novice contributor's ability to understand the framing and quality of video, but these lessons do not necessarily transfer to the capture of audio. Future work aims to investigate how audio capture can be structured and informed; developing template parameters analogous to those we have used for video as a way of describing audio qualities and capture techniques to contributors. Even given the success of templating for instruction and coordination in this context, questions were raised around the delivery of continuity from footage. In particular, the contributors' communication outside the system to manage who was shooting when, and maintain coverage was an example of how this form of template description is lacking in fully defining and communicating the constraints required to shoot some scenarios. We envision that the workflow flexibility provided through templating shoot requirements offer a platform for supporting new forms of participatory production, such as interactive documentary [1] or branching narrative [3] production which can fully leverage the rich meta-data and structure of content commissioned in this way.

#### **CONCLUSION**

Our aim in developing a template based commissioning framework was to provide a solution that was appropriable by both community and professional video makers. We did this by embodying a shared language which supported both clear instruction and a level of creative freedom within a digital tool. An important question for this deployment was whether non-professionals could work balance these two factors, and our results are very promising in this regard. However there are still important questions to be explored; in particular how professional broadcasters might work collaboratively within this space. We worked closely with our professional broadcasting partners in this deployment and we are confident that the quality of footage captured and the fundamental template approach to commissioning produces video which is up to professional standard and achievable without the overshooting that creates the editorial overhead typically associated with crowdsourced content. Future plans include our professional partners trailing the system in a full scale production where they take ownership of the commission platform.

#### **ACKNOWLEDGEMENTS**

*This project was funded by EPSRC award number SiDE EP/G066019/1 (Social Inclusion through the Digital Economy) and EPSRC award number DERC EP/M023001/1 (Digital Economy Research Centre). Data supporting this publication is openly available under an 'Open Data Commons Open Database License'. Additional metadata are available at: 10.17634/141304-2. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.*

## REFERENCES

1. Andre Almeida and Heitor Alvelos. 2010. *Interactive Storytelling*. Springer Berlin Heidelberg, Berlin, Heidelberg. <http://doi.org/10.1007/978-3-642-16638-9>
2. Daniel Arijon. 1991. *Grammar of the Film Language*. Silman-James Press. Retrieved September 7, 2015 from [https://books.google.co.uk/books/about/Grammar\\_of\\_the\\_Film\\_Language.html?id=6bQlAQAAIAAJ&pgis=1](https://books.google.co.uk/books/about/Grammar_of_the_Film_Language.html?id=6bQlAQAAIAAJ&pgis=1)
3. Tom Bartindale, Guy Schofield, Clara Crivellaro, and Peter Wright. 2016. TryFilm: Situated Support for Interactive Media Productions. *Computer Supported Cooperative Work*, ACM Press.
4. BBC. 2015. *Technical Standards for Delivery of Television Programmes to UK Broadcasters*. Retrieved from <http://www.bbc.co.uk/guidelines/dq/contents/television.shtml>
5. Dick C. A. Bulterman, Pablo Cesar, and Rodrigo Laiola Guimarães. 2013. Socially-aware multimedia authoring. *ACM Transactions on Multimedia Computing, Communications, and Applications* 9, 1s: 1–23. <http://doi.org/10.1145/2491893>
6. Renan G. Cattelan, Cesar Teixeira, Rudinei Goularte, and Maria Da Graça C. Pimentel. 2008. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Transactions on Multimedia Computing, Communications, and Applications* 4, 4: 1–24. Retrieved August 1, 2012 from <http://dl.acm.org/citation.cfm?id=1412196.1412201>
7. Gaby David. 2010. Camera phone images, videos and live streaming: a contemporary visual trend. *Visual Studies* 25, 1: 89–98. <http://doi.org/10.1080/14725861003607017>
8. Arvid Engström, Martin Esbjörnsson, and Oskar Juhlin. 2008. Mobile collaborative live video mixing. *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services - MobileHCI '08*, ACM Press, 157. <http://doi.org/10.1145/1409240.1409258>
9. Arvid Engström, Mark Perry, and Oskar Juhlin. 2012. Amateur vision and recreational orientation: CSCW, ACM Press, 651. Retrieved August 3, 2012 from <http://dl.acm.org/citation.cfm?id=2145204.2145304>
10. Martin D. Flintham, Raphael Velt, Max L. Wilson, et al. 2015. Run Spot Run. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, 747–756. <http://doi.org/10.1145/2702123.2702463>
11. David Philip Green, Simon J. Bowen, Christopher Newell, et al. 2015. Beyond Participatory Production. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, 3157–3166. <http://doi.org/10.1145/2702123.2702203>
12. Lyndon Kennedy and Mor Naaman. 2009. Less talk, more rock. *Proceedings of the 18th international conference on World wide web - WWW '09*, ACM Press, 311. <http://doi.org/10.1145/1526709.1526752>
13. Rodrigo Laiola Guimarães, Pablo Cesar, Dick C.A. Bulterman, Vilmos Zsombori, and Ian Kegel. 2011. Creating personalized memories from social events. *Proceedings of the 19th ACM international conference on Multimedia - MM '11*, ACM Press, 303. <http://doi.org/10.1145/2072298.2072339>
14. Mark Perry, Oskar Juhlin, Mattias Esbjörnsson, and Arvid Engström. 2009. Lean collaboration through video gestures. *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, ACM Press, 2279. <http://doi.org/10.1145/1518701.1519051>
15. Marco de Sá, David A. Shamma, and Elizabeth F. Churchill. 2013. Live mobile collaboration for video production: design, guidelines, and requirements. *Personal and Ubiquitous Computing* 18, 3: 693–707. <http://doi.org/10.1007/s00779-013-0700-0>
16. Guy Schofield, Tom Bartindale, and Peter Wright. 2015. Bootlegger: Turning Fans into Film Crew. *CHI*, ACM Press (Association for Computing Machinery).
17. Cees G.M. Snoek, Bauke Freiburg, Johan Oomen, and Roeland Ordelman. 2010. Crowdsourcing rock n' roll multimedia retrieval. *Proceedings of the international conference on Multimedia - MM '10*, ACM Press, 1535. <http://doi.org/10.1145/1873951.1874278>
18. Kai Su, Mor Naaman, Avadhut Gurjar, Mohsin Patel, and Daniel P. W. Ellis. 2012. Making a scene. *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval - ICMR '12*, ACM Press, 1. <http://doi.org/10.1145/2324796.2324829>
19. Raphael Velt, Steve Benford, Stuart Reeves, Michael Evans, Maxine Glancy, and Phil Stenton. 2015. Towards an Extended Festival Viewing Experience. *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15*, ACM Press, 53–62. <http://doi.org/10.1145/2745197.2745206>
20. Alexandra Weilenmann, Roger Säljö, and Arvid Engström. 2013. Mobile video literacy: negotiating the use of a new visual technology. *Personal and*

*Ubiquitous Computing* 18, 3: 737–752.  
<http://doi.org/10.1007/s00779-013-0703-x>

21. Fan Footage. Retrieved September 18, 2015 from <https://fanfootage.com/>
22. Burst. Retrieved September 18, 2015 from <https://www.burst.com/>
23. Seenit – Home. Retrieved September 18, 2015 from <http://seenit.io/>
24. BBC Taster - Ramadan in a day. Retrieved September 25, 2015 from <http://www.bbc.co.uk/taster/projects/ramadan-in-a-day>
25. The Five C's of Cinematography: Motion Picture Filming Techniques: Joseph V. Mascelli: 9781879505414: Amazon.com: Books. Retrieved September 4, 2015 from <http://www.amazon.com/The-Five-Cs-Cinematography-Techniques/dp/187950541X>
26. Bootlegger Open Source Software on Git Hub. Retrieved from <https://github.com/digitalinteraction/bootlegger-server>