Student Award Paper
Honorable Mention

Kerry Addica*

Descriptive Metadata: An Analysis of British Pathé Newsreel Collections from World War Two

DOI 10.1515/libri-2016-0085
Received July 1, 2016; accepted November 23, 2016

Abstract: British Pathé is considered to be the finest newsreel archive in existence, a treasure trove unrivalled in historical and cultural significance, containing 85,000 short film clips totalling 3,500 hours of footage (predominantly shot in grainy black and white with later works in colour) of major events, notable figures, fashion, travel, sports and culture, and is particularly strong in its coverage of the First and Second World Wars. In 2002, the process began to convert the film negatives into high-resolution digital files and upload them onto a computer database. A digital asset management system was utilized to host the archive and provide search abilities. One of the initial problems when converting the film negatives to digital files was a lack of detailed information. To combat this, the adoption of a better descriptive metadata scheme, using clearer, simpler keywords would make the site more user-friendly and easier to access. By comparing, sorting and categorizing a limited number of archival footage files from World War II on the British Pathé website, specifically “D-Day: 70th Anniversary Collection,” to create a statistical analysis of current descriptive metadata, a picture will emerge as to the user-friendly environment of the website. By using graphs, diagrams and stills, a deduction can be made as to whether the current keywords are well served or need improvement.

Keywords: DAM, metadata, archives

Introduction

Founded in 1896 by Charles Pathé, a pioneer of moving pictures in the silent era, British Pathé was a dominant feature of the UK cinema experience prior to the advent of television, broadcast news and social media, renowned for first-class reporting and informative style and credited with introducing the British public to the cinema newsreel (www.britishpathe.com). Extremely important in times of crisis, especially during World War II, it was utilized as a way for civilians to garner news about loved ones fighting on the European front. With bated breath, millions flocked to their local movie theatres for their weekly dose of news, worried whether they were going to catch a glimpse of a brother, husband, son, wife or daughter, hoping that the experience would be one of joy and not sorrow. In providing this service, Pathé News influenced a whole generation of Britons.

British Pathé is considered to be the finest newsreel archive in existence, a treasure trove unrivalled in historical and cultural significance, containing 85,000 short film clips totalling 3,500 hours of footage (predominantly shot in grainy black and white with later works in colour) of major events, notable figures, fashion, travel, sports and culture, and is particularly strong in its coverage of the First and Second World Wars.

Following its Golden Age from 1933 to 1958, and as televisions became a staple in every home, the newsreels became an antique, no longer necessary in a modern world with access to the news from the comfort of people’s living rooms, and production ceased in 1976.

Initially, the library contained a 50,000-card index with such categories as location, sport, transport and people. The card index was also available on microfiche and for many years the footage was only available to producers who licensed the archival newsreels for television and film productions. As technology advanced, it became apparent that greater access by the general public to the library would be favourable since the contents contain invaluable historical records. Therefore, in 2002, partially funded by the UK National Lottery, the process began to convert the film negatives into high-resolution digital files and upload them onto a computer database. A digital asset management system was utilized to host the archive and provide search abilities.

*Corresponding author: Kerry Addica, Department of School of Information, San José State University, San José, CA, USA, E-mail: kaddi002@gmail.com
Upon completion of the film negative transfer to digital, Pathé decided to make the materials available to the public via the Internet in the form of streaming video, utilizing Windows Media Player, on a dedicated website. This allowed the public to access films from the privacy of their own homes on personal computers free of charge, and with the popularity of genealogy on the rise, individuals could conduct research to locate relatives who had served in the war or find persons of interest or even themselves within the newsreels. With this in mind, an e-commerce component was added to the website to handle purchasing options; the 85,000 historical clips can be bought for personal use along with 12 million still images. DVDs can also be bought on the website, which features a link to Amazon for transactions and an additional add-on video subscription for a small monthly fee (much like an add-on subscription for HBO or other cable networks), with such titles as “Keep Calm and Carry On,” a three-disc set of original newsreels about war-related activities and events occurring on the British Home Front during World War II (www.britishpathe.com).

In 2012, Pathé won the FOCAL International Award for Footage Library of the Year. FOCAL is the Federation of Commercial Audiovisual Libraries, an international professional not-for-profit trade association formed in 1985 whose purpose is to facilitate the use of library footage, images, stills and audio in all forms of media production as well as promote and educate the need to preserve and restore footage and content.¹

In 2014, Pathé launched a dedicated YouTube channel comprising their entire collection in a concerted drive to make the archive more accessible to worldwide viewers (www.youtube.com/user/britishpathe). The channel has a new theme every month telling the story of a particular topic through archival footage. The current one is on India, entitled “The Golden Bird.” Past months have explored the story of the Republic of Ireland, lawbreakers and crime, and the motion picture industry. There is also a segment entitled “Newsreels of the week.”

Statement of Problem

“Perhaps more than any other media, digital information requires detailed metadata to ensure its preservation and accessibility for future generations” (Preservation Metadata for Digital Objects 2001, 4). Metadata is the structured knowledge that we have about things we collect and, in order to make a collection usable, we need to record what we know about the things in the collection and make that knowledge available to others. The record of our knowledge about our collections is metadata. We use that knowledge, that metadata, to manage our collections of things, to reveal them to others and to make them usable (Best Practices for Descriptive Metadata 2008).

Descriptive metadata record the attributes of a resource; they provide both intellectual access to content and access points by which users can discover digital materials, and can include elements such as title and author. Questions to ask when initially creating a record can include:

- What is it?
- Who made it?
- When was it made (date/year)?
- What is it about?
- Who is the intended audience/user?

One of the initial problems that occurred when converting the film negatives to digital files was a lack of detailed information. According to the British Pathé website,² the “clip descriptions are based largely on notes written by the cameramen and editors at the time they were filmed. This information wasn’t always accurate and some details are missing.” These quickly scribbled notes consisted of rudimentary information such as location, key people and what was happening within the shot. Since most of the footage was captured

during times of crisis, especially during World War II, recording the data to identify different locales, military units, specific battles, etc., was not a priority. Many reels of negative film contain different images from different battles all rolled into one, without a customary slate to differentiate one series of images from another (Austerberry 2006, 308). Therefore, when it came down to cataloguing the library and creating the record, the cataloguers’ job was a daunting one. Even so, all of this original information formed a valuable source for adding some descriptive metadata to the clips including, in some cases, a title, description, year, film canister number and whether the clip contains sound. Since some of the original information is limited in scope, or in some cases non-existent, the difficulty in adding descriptive metadata to these short films that contain rare yet nonetheless extremely valuable footage can be challenging if not impossible.

To combat this lack of information, while keeping in mind that the goal is to make the archive more accessible to viewers, the adoption of a better descriptive metadata scheme using clearer or simpler keywords would make it much more user-friendly and easier to access especially by a person who has limited computer literacy. The big picture here is that some of these users could be veterans themselves who are new to technology.

To create a standard as to how to catalogue and create descriptive metadata for such archival films could segue into an international standard that could be added to footage of other wars including Vietnam, Korea, and current footage of the conflicts in Iraq, Afghanistan and Syria. The standard could also be utilized for photographs dating back to the same era or war. Location and date, including month and year if possible, appear to be the most important information needed to access specific footage. Therefore, it would be advantageous to not just search by decade but by year, month and location.

**Methodology**

By comparing, sorting and categorizing a limited number of archival footage files from World War II on the Pathé website, specifically from the “D-Day: 70th Anniversary Collection,” to create a statistical analysis of current descriptive metadata, frequently used words and descriptions, a picture will emerge as to the user-friendly environment of the website. By using graphs, diagrams and stills, a deduction can be made as to whether the current keywords are well served or need improvement.

**Literature Review**

Of the different schemas or sets of fields that can be used to make up a metadata record, the Dublin Core Metadata Element Set (DCMES) is perhaps the most widely used. DCMES includes 15 core elements, plus a number of refinements (Dublin Core Metadata Initiative). Developed to describe born digital documents, DCMES was designed to sufficiently record the most basic units of information needed to facilitate basic end-user tasks (Coyle 2005).

Without metadata, an individual would not be able to use a search engine to locate resources, because if no metadata were attached to those resources, then nothing could be found. All databases containing digitized archives would exist in a void: they would simply be on the Web while being completely inaccessible to the general public. In fact, all e-commerce businesses would find it difficult, if not impossible, to turn a profit if they did not use metadata to assign keywords to their products. We need metadata, but how hard is it to assign those keywords to the resources? It all starts at the beginning of the process and complications can arise when the metadata are initially created and assigned to a resource. This stage is the most important, and according to the Getty Research Institute’s *Introduction to Metadata*, “quality metadata creation is [...] as important as the care, preservation, display, and dissemination of collections” (Baca 2016). In a museum setting, care is taken to preserve artefacts for future generations, lots of hours go into the preservation of the artefact and the same attention to detail should extend to the digital representation of those artefacts.

Let us consider a photograph taken during World War II. An individual, or a committee, must not only decide on the contents of the image but also predict what keywords a user may utilize in order to locate that image. To begin with, a word or piece of data, known as a “place holder,” should be the initial metadata added to a record. Then, as the resource moves through the institution, new departments or new personnel can add to this and make the metadata more specific to the resource so
that it is easier to locate for an individual searching a database. This is not a static process but an evolving one, because as new information is learned, the metadata relating to that specific item need to be updated on a continual basis.

It is a costly, time-consuming endeavour that requires specially trained staff, so an institution’s leaders need to comprehend its importance within the framework of a business model. If products are online in a database, or on a website, for the general public to view, then the keywords initially created for the resource should be an integral consideration in the process of conducting business. One of the successful things about the British Pathé website is its dedicated shop, “an e-commerce component to handle purchasing transactions. The public uses the site via a search engine. They can type a keyword or keywords, and the engine returns a list of all clips that are referenced by those words” (Austerberry 2006, 336). To add descriptive metadata to the newsreels of the D-Day landings would enable the end user to locate footage.

Since we now have a more defined idea regarding the need for descriptive metadata, it is vital to understand the background of the actual newsreel footage specific to this article to be able to decipher which beach and which Allied forces are included in the short clips.

The term D-Day relates to the Normandy landings, code-named Operation Neptune, of the Allied Invasion of Normandy, France, in Operation Overlord, on Tuesday, June 6, 1944. General Dwight D. Eisenhower, the Supreme Commander of the Allied Expeditionary Force, is quoted as saying, “people of western Europe, the hour of your liberation is approaching. All patriots, men and women, young and old, have a part to play in the final victory. This landing is but the opening phase of the campaign.”

The storming of the beaches was the “culmination of much preparation and represent[ed] the largest amphibious operation in the history of war” (Kindersley 2009, 258). The amphibious invasions began at 6:30 am and included 156,000 American, British and Canadian forces, whose goal was to capture and establish a bridgehead at five beaches, code-named Gold, Juno, Sword, Utah and Omaha. At the end of the day, 4,000 Allied troops lay dead, with another thousand or more wounded or missing. The battle for Normandy was the last great set-piece battle of the Western world and between June and August 1944, more than a million men fought one another in a last-ditch effort to decide the fate of Europe. If the Germans succeeded in pushing the Allies back into the sea, Europe would be ruled by Adolf Hitler and his Third Reich, but if the Allies got a foothold on French soil, they could make a concerted push for Berlin and force Germany to at least seek an armistice (Ford and Zaloga 2009, 7).

As the troops crossed the choppy icy waters of the English Channel as part of a flotilla of 5,000 ships and neared the Normandy coast, accompanied by 11,000 aircraft and 13,000 paratroopers who were dropped behind enemy lines, a barrage of never-ending bullets from camouflaged German pill boxes nestled in the cliffs hit the water, causing tiny splashes. When the doors opened on the landing crafts, the men carrying 80 pound packs waded through the sea to the distant beaches, praying to make it there alive. Whether they understood the magnitude of the task ahead of them is unknown; they were cannon fodder, a necessary sacrifice in the name of freedom and democracy.

Now that we know more about the historical background, context and importance of this footage, and have a greater understanding of what descriptive metadata are, let us take a look at some scholarly work that already exists to learn how other institutions have handled the same dilemma.

According to “Best Practices for Descriptive Metadata” (2008), one of the important reasons for creating descriptive metadata is to facilitate the discovery of relevant information. In addition to resource discovery by means of relevant criteria, descriptive metadata can also help to identify and bring similar resources together, distinguish dissimilar resources and give specific location information. Metadata can help to organize electronic resources, facilitate interoperability, provide digital identification and support archiving and preservation (“Best Practices for Descriptive Metadata” 2008, 1). Archiving and preservation are key terms for British Pathé, as the contents of the archive need to be handled carefully. Mauney and Zhang (2013, 2) discuss the fact that “an archival record, by nature, is not meant to exist on its own. The message it carries, the action it entails, and the consequence it incurs can only become fully comprehensible when the record has been associated with the [...] social context responsible for its creation [...] and other records to which it relates.”

---

In 2011, Binghamton University Libraries developed a digital preservation system for their special collections and archives department that consisted of 350,000 digital images. As part of the implementation, they extracted embedded metadata from digital photographs and transformed them into descriptive metadata (Corrado and Jaffe 2014). To make things as simple as possible, they used DCMES and controlled vocabularies like Library of Congress Subject Headings (LCSH) or Dublin Core Metadata Initiative (DCMI) Type Vocabulary to ensure the quality of metadata across the collection (Corrado and Jaffe 2014). Basic information taken from technical data written down by the photographer helped in inputting the descriptive information to create keywords such as location and photographic technical data from original photographs like shutter speed, film speed, time of day, etc., that could be linked to controlled subject terms to make it easier for an end user to access photographs. This same idea could be utilized for the British Pathé newsreels. The technical data regarding the newsreels already exist and just need to be added to the footage.

Data Analysis

Since the central question relates to the accessibility of the archive and user experience – especially taking into consideration an individual who may have limited computer literacy – focusing on the statistical analysis of the descriptive metadata currently present in a subset of the newsreels, specifically “D-Day: Beach Landing” (1944), which consists of 11 short film clips within the broader category of “D-Day 70th Anniversary Collection,” will lead us towards an answer. It is important to understand what descriptive metadata we currently have and deduce how well they work to be able to make suggestions on how to improve them. By comparing, sorting and categorizing frequently used words and descriptions, as well as the use of graphs, diagrams and film stills, a picture will emerge of how user friendly the current environment is.

To begin with, let us explore the current general metadata each newsreel contains to understand if they are consistent. The database fields are shown in Table 1.

Table 1: Current metadata for “D-Day: Beach Landing” newsreels.

<table>
<thead>
<tr>
<th>Subject metadata</th>
<th>Film metadata</th>
<th>Tape metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Canister number</td>
<td>Film ID /tape number</td>
</tr>
<tr>
<td>Summary</td>
<td>Duration</td>
<td>Time code in</td>
</tr>
<tr>
<td>Description</td>
<td>Length in feet</td>
<td>Time code out</td>
</tr>
<tr>
<td>Issue date</td>
<td>Sound/silent</td>
<td>Cameraman</td>
</tr>
<tr>
<td>Year and decade</td>
<td>Music cue</td>
<td>Slate/clipboard</td>
</tr>
<tr>
<td>Keywords</td>
<td>information</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Film stills</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Film stock /black</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&amp; white or colour</td>
<td></td>
</tr>
</tbody>
</table>

These metadata fields are broad in nature, encompassing everything needed to provide accurate information and further provide adequate material to create specific keywords. They should also provide a variety of newsreel results for each new search.

The 11 short film clips relating to the D-Day beach landings contain all of the descriptive metadata fields from Table 1 but the specific data are absent. The dates and location are unclear or unknown, there is no sound, and the data from the clapboards/slates embedded within the footage are absent. Table 2 describes the extent of the lack of information.

Table 2: Incomplete information for “D-Day: Beach Landing” newsreels.

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Dates unknown</th>
<th>Location unknown or unclear</th>
<th>Unused material</th>
<th>Unissued material</th>
<th>No sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of newsreels out of a total of 11</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

The main problem is lack of information relating to location, followed by a lack of dates. To ascertain if these data can be found and added to each newsreel, each film was analysed to see which ones contained embedded clapboards/slates and how much useful data were on each one. Additionally, a comments field is provided at the bottom of each newsreel for the public to add their ideas as to the location of the footage, with a disclaimer stating: We always welcome comments and more information about our films. All posts are reactively checked. Libellous and abusive comments are forbidden.

Of the 11 newsreels, four newsreels had comments and two had clapboards/slates embedded within the footage. One containing an embedded slate was picked and viewed.

---

8 For Library of Congress Subject Headings, see http://id.loc.gov/authorities/subjects.html (accessed March 20, 2016).
for further analysis: Beach Landing (2 minutes, 41 seconds) is silent grainy black-and-white footage of soldiers, heads bowed, ducking bullets, transported in a landing craft. When the door opens, an endless succession of soldiers exit into waist-deep water weighed down by heavy packs, rifles in hand, the weapons held above their heads to prevent them from getting wet, the craft bobbing in the water, as they clumsily, but in an orderly manner, wade out towards a distant foreboding beach, barbed wire barriers dotting the sand, burnt-out two-story houses, crumbling ruins, sitting on the beach’s edge. A cataloguer’s note reads: Very possibly French beach but after the initial D-day landings. Description reads: Unused /unissued material – dates and locations unclear or unknown. Beach landing. LS of soldier on beach. CU from inside landing craft as British soldiers jump out and wade through water towards the shore. Some good footage.

Viewing the clip elicited more specific information, currently not part of the metadata, including the following in the embedded slate (Figure 2):

- O.S.S.
- U.S.N.
- ARMISTEAD

A general Google search of Mark Armistead OSS returned an excerpt from a book entitled John Ford: A Man and His Films by Tag Gallagher (1986), which states that “later on D-Day, Ford joined Armistead in a PT boat” (Gallagher 1986, 217). The Ford in question is the Hollywood movie director John Ford, best known for cinematic classics such as The Grapes of Wrath, How Green Was My Valley and The Searchers, arguably the best Western ever made. Ford is quoted as saying, “You say someone’s called me the greatest poet of the Western saga. I am not a poet, and I don’t know what a Western saga is” (McBride 2001, epigraph)

According to the book Searching for John Ford by Joseph McBride (2001), Ford reported for active duty in 1941 to the chief of naval operations and “formed the Field Photographic Branch of the OSS to document military activity [...] and provide mapping and other photographic reconnaissance information” (McBride 2001, 275). On D-Day, he was placed in charge of cinema photography (McBride 2001, 394) and had the idea of “installing fixed cameras and camera mounts aboard landing craft that were to make the initial assault. Ford utilized 152 cameras set to automatically begin operating as soldiers left the landing craft” (McBride 2001, 395). Five hundred 35-mm cameras, each loaded with four minutes of film, were mounted onto landing craft and the resulting coverage “turned out to be the most successful and [...] best film shot during the invasion of the beaches” (McBride 2001, 395). Ford credited Lieutenant Mark Armistead, USNR, OSS, and his men for covering the entire embarkation coast and “working 24 hours a day for six days” (McBride 2001, 395).

Mark Armistead’s real name was Marcus E. “Mark” Armistead and he had been working at a Hollywood equipment rental house when Ford first met him. When the war broke out, he signed up for Ford’s Naval Reserve unit and in 1944 was posted to London working as Ford’s aide and was one of the cameramen sent in on D-Day (McBride 2001, 346). After the war ended, Armistead fondly remembered Ford as “the only Hollywood director that fought who did not forget his men” (McBride 2001, 335).
So, what is the OSS? On the website of the Central Intelligence Agency (CIA),\textsuperscript{10} an article entitled “What was OSS?” explains that the Office of Strategic Services was the precursor to the CIA, established in 1942 as an intelligence-gathering branch of the government formed to coordinate the behind-enemy-lines spying activities of branches of the United States Armed Forces.

Both pieces of information, Armistead and OSS, would be beneficial information to be added to this newsreel, since both go hand in hand and were integral to the operations of D-Day.

Another source of information for these vintage newsreels could be found at the British Pathé War Archives’ YouTube channel, which contains 85,000 historical films and has 64,413 subscribers (as of April 26, 2016). A general search for D-Day Beach Landing 1944 returns a single result in the form of a 45-second film entitled D-Day Beach Landing (1944),\textsuperscript{11} which has 28,888 views. The footage is of brave soldiers exiting a landing craft into the ocean with a beach in the distance, much the same footage as the newsreel “Beach Landing” from the British Pathé website. The only difference is the YouTube video has added post-sound design of bombs dropping, bullets flying and the bloodcurdling sounds of war.

Scrolling down to the comments elicits help with additional information relating to the contents of the video; the most interesting is a debate on what country the soldiers came from. The description that accompanies the video explains that they are British but the general consensus of the comments is that they are in fact Canadian troops and the landing beach is a toss up between Sword and Juno. One of the commenters explains that you can tell which troops they are from their helmets, insignias, uniform and vehicles.

Utilizing the knowledge of the general public in identifying key elements of the unknown location and dates of the footage, including who the cameraman could be and the beach the soldiers are landing on as well as what country they are from, can be beneficial to accurately add additional descriptive metadata to the footage of the D-Day landings located on the British Pathé website.

Adding these data to the newsreels can facilitate access and make the database and website much more user friendly. Since we could potentially add the country of the soldiers within each separate reel, a user could search for Canadians, British, Americans (US) and others. It could help narrow down the search specifically.

### Results and Discussion

Since the central question relates to the user-friendly nature of the British Pathé website and whether the material in the database is easily accessible to a person with basic computer ability, and whether or not the current metadata assigned to the footage are advantageous to that end, it appears that with further research into each individual film, new information can be located which, if added to each individual film, would alleviate the problems of current metadata fields, such as unknown location and dates, and would aid an individual in locating specific footage.

The problem is that this would require a tremendous amount of additional manpower, not to mention expense. To find the additional information would require not only an analysis of each individual film clip but would also require cross-referencing with British Pathé’s YouTube account as well as the Imperial War Museums’ archives.

The information gained would then have to be input into the currently absent fields. Of course, volunteers or interns could be utilized to complete this work at a cost of zero cash outlay and it would be an exciting as well as an educational experience for the intern or volunteer. British Pathé could begin a programme specifically catering to


this idea and link to universities, especially those that offer graduate degrees in Library and Information Science.

The next question is whether or not this additional research for each individual film clip is necessary to fill in the gaps of absent information. In terms of historical significance, it is extremely important. To have accurate information on this rare footage would be useful to both scholars, students of D-Day, relatives of deceased soldiers, and veterans alike. This rare footage is part of our collective history and is of human interest.

A central question appears to be whether or not this archive is for the benefit of humanity or just to make money. In terms of mankind, as World War II becomes a distant memory for many of the world’s current population, it is always necessary to remember where we came from, the wars we fought and why we fought them. For many people who can imagine a world where the Allies lost the war and Adolf Hitler and the Third Reich ruled Europe, this footage, especially from the “D-Day: 70th Anniversary Collection,” should be treated as a global treasure trove of invaluable information.

Conclusion

British Pathé newsreels provide a porthole into the past, encompassing a database containing over 85,000 short film clips totalling 3,500 hours of film. Initially, the library contained a card index that was converted into digital files. A digital asset management system was utilized to host the archive and provide search abilities.

Upon completion of the film negative transfer to digital, Pathé decided to make the materials available to the public via the Internet in the form of streaming video on a dedicated website, but there was a problem: lack of information to input accurate metadata fields to facilitate easy access and make the site as user-friendly as possible, especially to a person who has limited computer literacy, especially veterans.

Research was needed to gauge the feasibility of attaining additional information for specific film clips. By comparing, sorting and categorizing a limited number of archival footage, specifically a small collection of 11 films from D-Day, a deduction was made and the answer was yes, it is possible, but patience is needed, as well as a sound methodology and the cross-referencing of photography stills to film footage, and utilizing other sites such as the Imperial War Museums’ archives.

Since many analogue and antiquated technologies are currently being digitized, one of the foremost accommodations needs to be the user-friendly nature of the archive. Descriptive metadata describe a resource for purposes such as discovery and identification. A reason to devote precious resources, including time and money, to complete this very important task, and to be as specific as possible with the information needed to create accurate and fully realized metadata, is because we want the public to access our treasures and we want their experience to be relevant in today’s political climate. We study history to understand where we are today in relative terms to where we were yesterday. Lest we forget, destined to remember.

References


