2016 Conference Transcription

Date	Friday
Session Title	Thriving in Uncertainty
Session Time	14:30 – 14:50
Moderator	n/a
Speakers	Ruth Garcia Gavilanes
Notes	n/a

Introduction

Voiceover	Hello and welcome to Future Everything's 2016 Festival Podcast Series. Over two days, in Manchester's iconic Town Hall, we tasked designers, artists, scientists, and many more, to rethink our resources. From life, earth and intelligent to community and uncertainty, our speakers asked what we might need less, and more of, in our near future. Doctor Ruth Garcia Gavilanes works as a computer social scientist at the Oxford Institute of the Internet. A previous researcher at Yahoo! Labs and Telefonica Research and Development in Barcelona, Ruth's research asks questions such as how does social media make us remember an event across the world and is the data actually as accurate as we think it is?
Speaker	So I'm going to talk about online collective memory, but before going into online collective memory, I'd like to talk a bit about collective memory, only the term collective memory. Probably this term was first tackled by a sociologist Émily Durkheim at the beginning of the nineties for the first time, although he never really used the term itself. But he noted that every society had a need to connect with the past, to have a sense of continuity with the past, and that this past was going to bring some sense of identity and cohesion to the societies. So the way to remember the past, or commemorate the past, was bringing the groups together and trying to remember what happened through rituals, symbols etc. It was his student actually, Maurice Halbwachs, that later used the term itself, collective memory. He believed that societies are the ones that dictate how we shape memories and how we remember. He didn't believe in individual memory. He believed in individual memory only in the context of groups, of societies.

	So probably, there are two examples that portray best what collective memory means. The first is the holocaust. It's a memory that we all share. There are museums around the world that remind us about narratives, the facts, and that also plays a key role in the identity of the Israeli community. The other example is the Hiroshima and Nagasaki memorials in Japan. This also makes a crucial part of the post-war Japanese identity and also reminds us of the dangers of nuclear weapons. But collective memories are not only retelling of unbiased truths. Actually, to form a collective memory, it requires a lot of discussion, a lot of arguments, fights. This is important, because once it is formed and is passed from ears and from generation to generation, collective memory can be at the root of nationalism, prejudice and sometimes war. The way collective memories are formed now, shaped now, might differ drastically from what came or have been before. It's that now, thanks to the internet, we can shape our memories, our history through online. The benefit, good or bad, these memories are not built only within a group anymore but now it's being shared. People from all around the world are making part in the formation of memories. Many sites, for example for 9/11, keep track of every revision, every argument that took place to form memories. The good thing is that we can collect this information, measure it and study. Certainly there are many ways to do that, but today I'm going to talk about one thing in specific and that is Wikipedia. And I'm sorry, [inaudible 04:38] but okay. When you think about the best repository of human knowledge, what do you think about? For many people, probably it is the British Library or the Library of Congress in the States. For many researchers, and me, it is Wikipedia. There are many reasons that we use Wikipedia to study collective memory or memories. One of these can be, that while it's the most up-to-date encyclopaedia nowadays online, generally when there's an event, a very [inaudible 05:16] event happening now, it takes sometimes hours, sometimes even minutes to release an article online. It's the sixth, I believe, most visited site on the internet. It has now around four million English articles that are enough to fill nine hundred and fifty volumes of the Encyclopaedia Britannica. So this is one of the reasons. But probably, most importantly, Wikipedia is a framework that lets us measure and see how all this knowledge, all these memories, took place; what was accepted, what was not, what was agreed and what was not. This is possible because every article has a story. It has the revision history of every edit in the article, recording and saving information about who added what and when and why. Besides that, every Wikipedia article has a talk page, where editors can actually agree, talk or fight sometimes about issues or subjects about articles. If we were to see where these editors come from, they come from all over the world, and they edit different articles in different languages in Wikipedia.

	Actually, there was one historian, a blogger, that decided in 2010 to print the whole revision history of the Iraq War article. He printed and he ended up with twelve volumes made of ten thousand pages, with twelve thousand editions, just for one article in Wikipedia. And every one of these revisions had a story. Sometimes there were people upset and putting some bad things, so later another editor would erase. In fact, there was one revision where they erased everything and they said Saddam Hussein is just a d***head and then there was another editor erasing it immediately. So all these revisions have a history and tell us exactly how this article came to light. If we were to see as well these additions, these edits, these revisions in the time when they were made, we could actually find some insights as well. For example, here we have the Chernobyl disaster time series for the revisions. The article came to light on the third of September of 2002. We can see that the revisions took place all through the year. In the X axis, for each label, we have the anniversary of the Chernobyl disaster. The time series goes from 2002 to 2015, and through all these years there were modifications in the articles with some spikes, one when a journal of cancer was released talking about the disaster, and the other one in 2011 when the Fukushima disaster happened. Another example is the September 11 attacks. It was created in November of 2001, the article, but additions happened and modifications happened all through the years, until 2015, with spikes specifically on its anniversaries and in 2006 when there was a new tower built, and in 2011, close to the anniversary, when Osama Bin Laden was killed. Finally, we have a different scenario, time series of editions of the London bombing attacks in 2005. Here it's quite different, this time series. Here we see that at the beginning, when it was created right on the same day, the editions happened mostly at the beginning and very few later. When we saw why this happened, we found out that many editors were fighting about little things about words, what kind of words they had to use. Terrorist, terrorism? So the majority of the edits actually happened at the beginning. So Wikipedia allows us to find this information about editors, but similarly, there's also the possibility of knowing how many people visit each article. If we were to see, for one day only in 2011, where these people come from, where all the views to the articles come from for the English Wikipedia, we'll see that they come not only from the English speaking countries, but also from parts of Brazil, from all Europe, even from, I think, India. If we were to also plot the time series of the same examples, but from the viewership data instead of the editions, we could also find some information, some insights. For example, for the Chernobyl disaster, the story is quite different from the editions. Something I forgot to mention, the viewership data in Wikipedia is available only after 2008. So this is viewership data after 2008. Here we see that actually people were not very interested in the Chernobyl disaster before 2011, but the Fukushima disaster happened, people started going to the article and they started reading about the Chernobyl disaster. For the September 11 attacks though, we see that people go back to the article every anniversary, with some spikes when Osama Bin Laden died. For the London bombings in 2005, also

	people go back in the anniversary every year, but there are some fluctuations around the anniversary and there is one spike when some verdicts came to light in 2011. From this time series, we can see that when something new happens that is related to the past, people go back and read old articles. My team and I were interested in this type of behaviour and what we did is we collected all the articles in English about airline crashes, and investigated this behaviour a little bit more. I will share with you some examples. This is the time series for the German Wings flight that probably we all remember. That happened in 2015, the crash that was on the Alps killing hundreds of people, due to a suicide attempt. The Wikipedia article was created on the same time. We can see there is an immediate, huge spike of attention at the beginning, and then it naturally falls back. Then there is another spike some months later, because new discoveries came to light. This is the time series only for the three months. If we were to plot this at the same time, the time series of all articles about all airline crashes, we have this. These are the time series of all the articles about airline crashes that happened before. We see that a lot of them have a spike, but the one having the biggest number of views, the most [inaudible 13:30] memory here was to the article about the Egypt Air flight that happened in 1999. Probably we all forgot about it by now, but this was the article that was more triggered. Why did this happen? Was it because there was a direct link, maybe, in the article? If we zoom in to the seven days, we will see that here we marked with blue, the ones that had a direct link, and with red, the ones that did not. And indeed we see that the Egypt Air flight had a direct link with the German Wings flight. But the second most triggered article that goes to the Pacific Southwest Airlines flight that happened in 1987 in the States, also a suicide attempt and killing around fifty people, was also triggered, but this one did not have a direct link. Why did this happen? Was it because of external sources? Is it because it was in the States? Is it because there was a link there and removed later? These are some questions that we are still tackling with my group. On top of that, we also were interested in comparing these curves with a German Wings curve, but not only how triggered or how high the views were triggered when there's a new event, but also how similar is the shape? How the attention goes along with the German Wings flight. We did that, but not only for the German Wings flight, but for every possible airline accident and we paired them, compared their time series, and grouped each pair according to the distance in time and distance in location. We have several groups. We made sure that these groups had the same amount of pairs. We found out that for the shape of the curve, and only considering those that have moderate to more correlation between them, we found that people generally would tend to follow the same attention for cases

	that are similar, recent in time, close in time and close in location. We're releasing a paper soon about this. So we have seen from these few examples, that with Wikipedia, we can not only see how memories are shaped, we can also measure how people remember and how people forget. I have worked a lot with social media sites like Twitter and so far Wikipedia's the one providing the most granulated data about memory. I will end with this quote and say that probably, for the first time, we don't know until when, maybe one day it stops, but now we have a system that allows us to track down all these little pieces of information about memory that challenge absolutist views of the past, and therefore absolutist views of the present, and future. Because of that, we need to find ways of giving this data information. We had this data there before for a lot of years, but it's just recently that we are starting to understand and measure and do something with it. It's important to do that so that we know how our history is being written now. Thank you very much.
Voiceover	We hope you enjoyed Ruth's talk and thanks for listening. You can hear the rest of the talks from 2016 at futureverything.org/2016podcasts.

[Transcription ends]