

## **Contemporary collecting: collecting Instagram for local studies.**

**Ellen Forsyth, Ngarie Macqueen and Daniel Nitsikopoulos**

### **Abstract**

When the James B. Hunt Jr. Library opened in 2013 they collected Instagram images using #myhuntlibrary and shared their code on Github.

In 2016 with the support of the State Library of NSW three public libraries started collecting local material from Instagram, using this code. The two libraries who continued with this project provided multiple hashtags. Additional coding has been done to make it easier for the libraries to tailor the hashtags to current community needs. The code is shared on Github so that others can do this work.

The session is aimed libraries who collect local studies or material about their community of users. They will be encouraged to consider collecting social media will see that this can be relatively low cost both in dollars and staff time and enables the collecting of new local material. This is a flexible framework which can be tailored to suit the community being served and has the potential for community involvement in suggesting the hashtags collected.

The large quantities of community created digital content highlight the importance of contemporary collecting for local studies. This method of collecting Instagram, while not perfect, is showing one way to consider the infinite possibilities of collecting and displaying recent material for local studies in public libraries.

### **Introduction**

Collecting social media is a topic which has been raised at the NSW Public Library local studies working group meetings for a few years, as it is an area where new content is being created but it continues to be a challenge to collect. Social media is a way to see how people are recording where they live, work or visit and this personal documentation can provide a useful perspective for public libraries to include as part of their local studies collections. The State Library of NSW #NewSelfWales<sup>1</sup> and Social media archive<sup>2</sup> show other approaches to this kind of collecting. This paper is to show a possible method for public libraries collecting social media for their local studies collections.

This is not a perfect tool which we are discussing, but perfect tools do not exist. It is important to explore possibilities in a constructive and practical way. This paper will provide the background to the collecting, a case study and a developer perspective. We will all speculate about possibilities for the future.

### **Background to project**

When the James B. Hunt Jr. Library opened in 2013 they collected photographs which people were taking of this new library and how they were using it. The James B. Hunt Jr. Library in Raleigh, North Carolina, was built as a way for North Carolina State University to encourage

---

<sup>1</sup> #NewSelfWales ux research + testing. (n.d.). Retrieved from <http://dxlab.sl.nsw.gov.au/blog/newselfwales-ux-research-testing> [accessed 13 August 2018]

<sup>2</sup> Live - SLNSW Social Media Archive. (n.d.). Retrieved from <https://socialmediaarchive.sl.nsw.gov.au/prototype/liveV4.html> [accessed 13 August 2018]

research and innovation. Library staff asked people to use the hashtag #myhuntlibrary<sup>3</sup> when they posted photographs of the library on Instagram. As well as collecting relevant social media the James B. Hunt Jr. Library produced a social media archive to assist other organisations in collecting social media<sup>4</sup> and they shared their code for collecting on Github<sup>5</sup> (see also Appendix 1). The images collected by the James B. Hunt Jr. Library showed that there were near infinite possibilities for collecting images based on searches of hashtags for local studies collections in public libraries.

Historically, libraries have made their collection (and therefore information) available through the use of a select number of pieces of information, such as subject headings, either created in-house by staff or taken directly from a standard source such as Libraries Australia holdings or the Library of Congress Subject Headings. With the introduction of the World Wide Web and the move to Web 2.0, information has become a more fluid, social and shared experience whereby individuals users can not only source information, but are empowered to create. This Web 2.0 environment has lead to users creating their own folksonomy; a collaborative way of classifying key information from the content being created, for not only their work, but also value adding to others creative work. In the many social media platforms, this is achieved through the use of hashtags - a series of words, numbers and symbols preceded by the symbol # that can categorise, describe and/or collate the content of the image, tweet or blog post that has been created.

Despite early interest from Public Library Services staff at the State Library of NSW, it was a few years before it was able to be tested with public libraries.

### **Public library involvement part one**

During the first half of 2016 approval from Manager, Public Library Services, State Library of NSW was provided for testing this Instagram collecting with one NSW public library. This project was able to take place through funds, coordination, and support provided by the State Library of NSW and the skills of the team at Icelab, a software development studio based in Melbourne and Canberra. A set of criteria for public library involvement was developed, and there was one library which matched this. Sutherland Library were approached directly. The library staff were keen and their work helped highlight the need for clearer expectations for this kind of project. It also showed the value of trying this method of collecting social media, and the need to test this methodology with more than one library. This was understandable given that this was a new tool to test and there were many unknowns. Finding out some key information, such as how much time needed to be spent on it, was one of the reasons for this trial. The results of this testing with one library looked promising, even though Sutherland Library decided to not continue with this trial. The work done by Sutherland Library as the first test site in NSW provided the impetus for further exploration of social media collecting by NSW public libraries. Fortunately there was able to be a part two to this project to build on what was learned and to further test the potential of this collecting.

---

<sup>3</sup> My #HuntLibrary - NCSU Libraries. (n.d.). Retrieved from <http://d.lib.ncsu.edu/myhuntlibrary> [accessed 9 August 2018]

<sup>4</sup> Social Media Archives Toolkit. (n.d.). Retrieved from <https://www.lib.ncsu.edu/social-media-archives-toolkit> [accessed 9 August 2018]

<sup>5</sup> NCSU-Libraries/lentil. (n.d.). Retrieved from <https://github.com/NCSU-Libraries/lentil> [accessed 10 August 2018]

## **Public library involvement part two**

The second phase of involvement was for two more libraries to test the social media collecting, building on the work of Sutherland Library. This involved interested NSW public library staff writing an expression of interest addressing feedback from the first library involved, as well as from staff from the State Library of NSW. This expression of interest (shown in Appendix 2) aimed to be clear about expectations and requirements. The expression of interest led to five expressions of interest. While three did not match all the criteria and included a library outside NSW, two did.

Once the participating libraries had been determined there was a video conference to connect the two libraries with each other as well as with the staff at Icelab. This use of technology allowed people in different locations, Balmain, Canberra, Casino and Sydney CBD to be able to meet together. This first video conference enabled a discussion of requirements which each library had to resolve within their own councils. The persistence of these libraries, Richmond Upper Clarence Regional Library and Inner West Libraries, in following through on this project led to two very interesting sets of images being collected. The libraries had to work with their council IT staff to provide the required information and access.

While #myhuntlibrary uses one hashtag, this project asked the libraries to consider which hashtags showed the many faces and environments of their community. The hashtags being collected can be edited by the library staff to include new topical areas. From the photographs which use the chosen hashtags, libraries select the images which are collected, choosing how many to add. This allows them to target particular areas of collecting.

As the number of hashtags increased it became harder for the library staff to work with them. Additional work was done by Icelab to make accessing the hashtags in use easier, rather than being combined with every hashtag used on the collected images. This update has been made available on GitHub<sup>6</sup> so that other libraries considering this kind of collecting can have access to these changes.

The following case study from Richmond-Upper Clarence Regional Library shows the experience of one of the libraries involved in this project.

### **Case study from Richmond Upper Clarence Regional Library**

The Richmond-Upper Clarence Regional Library (RUCRL) began in 1978, with the library network currently supporting the information needs of the 33 000 residents and visitors to the Kyogle and Richmond Valley Council (RVC) local government areas in the far north coast of New South Wales. The library network has no separate local history policy, no photography collection, no digitisation plan, and the library's Instagram account was created in February 2016. This meant that the library service wasn't locked into an existing framework for local history collecting and were therefore free to explore new approaches to how the library service collects information about the community it is a part of.

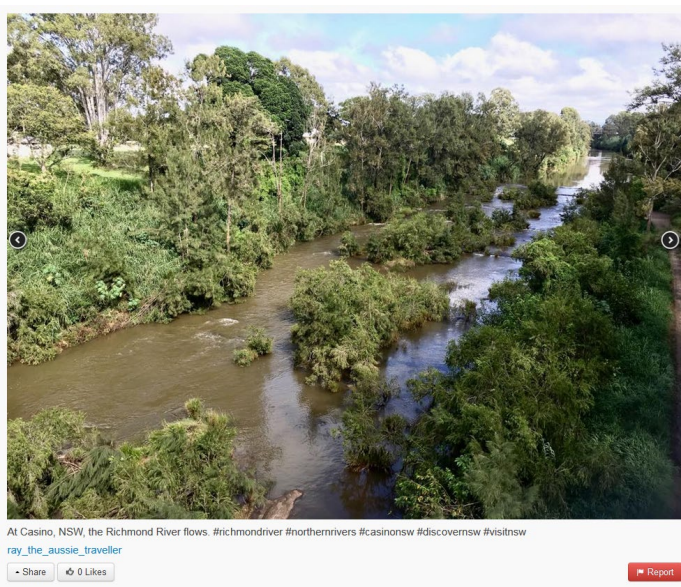
Once the library had been notified that they were eligible to participate in the second phase of the trial, it required open communication and support between myself (Ngarie who is the RUCRL contact), the RUCRL Manager Regional Library, key RVC IT staff, key staff from Icelab (Daniel), and Ellen from the State Library of NSW. The trial was successful due to all individuals involved being proactive, responsive and communicative. This was crucial due to

---

<sup>6</sup> slnsw/lentil. (n.d.). Retrieved from <http://github.com/slnsw/lentil> [accessed 9 August 2018]

the logistics of setting up the various accounts and monthly payments required, choosing and registering the domain for the approved images to be showcased, training in how to use the Lentic (Instagram photo harvesting) software and the linchpin to what determined what collection would eventually contain - the selection of the hashtags used to harvest images from Instagram. The list of hashtags decided upon were based on the towns and locations, the biggest festivals held and the local Aboriginal people, Bundjalung (as shown in Appendix 3).

This list was compiled based on the many ways people use hashtags in social media. Hashtags can be a simple description of the content, as shown is this picture of the Richmond River, which uses the hashtags: #richmondriver #northernrivers #casinonsw #discovernsw #visitnsw<sup>7</sup>:



---

<sup>7</sup> Ray on Instagram: 'At Casino, NSW, the Richmond River flows. #richmondriver #northernrivers #casinonsw #discovernsw #visitnsw #seeaustralia #travellaustralia?? (n.d.). Retrieved from [https://www.instagram.com/p/Bgr2jU2n1WR/?taken-by=ray\\_the\\_aussie\\_traveller](https://www.instagram.com/p/Bgr2jU2n1WR/?taken-by=ray_the_aussie_traveller) [accessed 11 October 2018]

Or they can use humour, as shown in this example:  
#coupleswhorenovatetogetherstaytogether<sup>8</sup>:



Or the user can use hashtags as a way to make their content more discoverable by connecting to other popular content. As seen in the example below which uses hashtags for Brisbane, Mount Tamborine, Gold Coast and Kyogle in an image taken in the main street of Kyogle<sup>9</sup>:

---

<sup>8</sup> Heavens Head on Instagram: "Babe it's all about the vision." He says. Me says "I'm squinting my eyes, but just can't see beyond the holes..." #trusthim ... ..?? (n.d.). Retrieved from [https://www.instagram.com/p/BXK2lF-BTJD/?taken-by=evans\\_head\\_reno](https://www.instagram.com/p/BXK2lF-BTJD/?taken-by=evans_head_reno) [accessed 11 October 2018]

<sup>9</sup> Sophie Baker on Instagram: 'Margot having her Sex and the City moment in the main street of Kyogle ? #onemoorewedding . . . . . #sexandthecity?? (n.d.). Retrieved from <https://www.instagram.com/p/BUQ5rtRgjXP/?taken-by=sophiebakerphoto> [accessed 11 October 2018]



Margot having her Sex and the City moment in the main street of Kyogle 🤪..... #sexandthecity  
#carribradshawmoment #kyogle #kyoglewedding #sophiebakerphotography #goldcoastphotographer  
#goldcoastweddingphotographer #goldcoastweddingphotography #mounttamborinephotographer  
#tamborinemountainphotographer #nikon #aipp #weddingphotographer #weddingphotography  
#brisbaneweddingphotographer #brisbaneweddingphotography

sophiebakerphoto

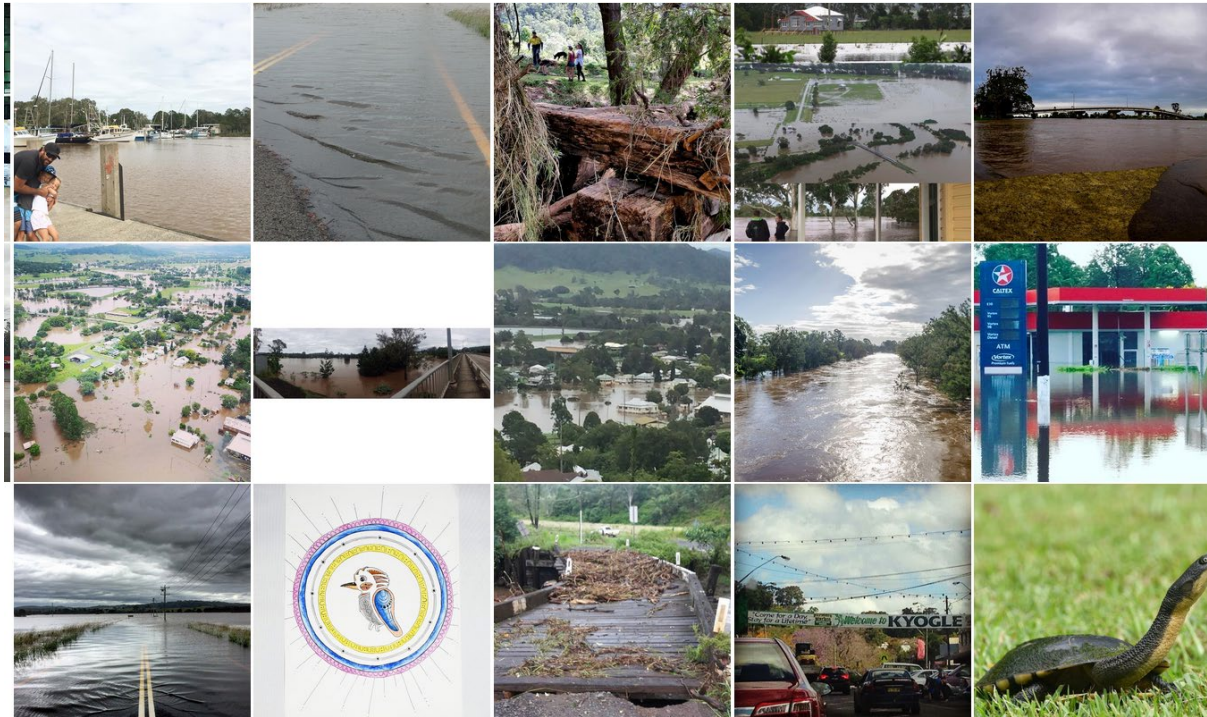
Share 0 Likes

Report

The next important step is how to sort through the thousands of images harvested to collate the specific images that will become a part of the Richmond Upper Clarence website - for a contemporary local history collection<sup>10</sup>. The RUCRL collection development policy aims to collect all materials possible relating to the communities served, so therefore a wide variety of images were selected for the website. These images cover areas such as: local geography, real estate, businesses, flora, fauna, festivals and events, food and drink, and Bundjalung art and celebrations. Images of both high and low quality were selected to demonstrate a cross section of what was being shared. A fifty-year flood occurred just as the project was started and so could be documented and displayed as it was happening.

---

<sup>10</sup> Richmond-Upper Clarence Regional Libraries - Richmond-Upper Clarence - Richmond-Upper Clarence Regional Libraries. (n.d.). Retrieved from <http://images.richmondvalley.nsw.gov.au/> [accessed 11 October 2018]



*Photograph 1 Showing screenshot of images collected about local floods*

This event alone proves that even if the process of harvesting images is only done over a short time frame, it will still enable future interested individuals to compare the then with the now. The RUCRL is on the wrong side of the digital divide with much of the region unable to access high quality, fast internet. Some areas have higher levels of individuals with a lower socioeconomic status, making them less likely to be able to afford the computers and mobile devices required to access social media platforms. In addition, most users of social media, including Instagram, are young adults. Rather than these being reasons for not collecting this type of material, it gives a specific context for who, and what was being shared at the time. Just as it was only those that could afford a camera in the past that could afford to take and print out physical photographs. With this project, the library service has had the privilege of being able to collect, collate, and display history as it's being made, from the very people living and creating it.

### **Case study for technology**

The Instagram harvesting tool is made up of a handful of small parts that work together to provide a relatively flexible tool for collecting and classifying images published online. These include a system for managing tags, harvesting images based on those tags and then moderating those images so they can be displayed to the public.

When choosing a tool to assist in archiving images from Instagram based on specific terms, it is helpful to keep a number of factors in mind. Ideally the software is able to provide some generalised workflow that is easy to adapt to, regardless of the end user's technical knowledge. In our case, Lentil uses common names for its internal workflows, so when a user is learning the software for the first time, terms like "tags" and "tag sets" don't seem totally alien.

The next thing that's important to consider is how easy it is to extract the archived images from the system itself. It is one thing to be able to log in to the application and look through

page after page of images, but there needs to be a way to extract the raw image files. In our case, all of the images and videos are backed up to Amazon's cloud storage service (S3) and can be downloaded as a single folder.

For the tool to have some benefit to those just wanting to browse the collection, there needs to be some form of curation otherwise the collection of images becomes unwieldy and tends to lack focus. This is possible with Lentil's approval process. Once an image is harvested, it must be reviewed by a person and allowed to be viewable through the public facing website. There are also tools in place for the public to report an image that they think is inappropriate or to comment on an image if they happen to have some more information about the subject.

In order to aid with discoverability, the Lentil system also adds tags that are extra to the chosen tags in to the system. For example, if the tag #mytag is set to be imported, and an image with the tag #mytag also has #myothertag on it, the system will add #myothertag to the list of tags that can be imported in the future. This is to help curators find new tags that might otherwise be related to tags they are already importing.

During the first few months of use, this was somewhat manageable, but as time went on, and the number of tags in the system grew, it soon meant that the method we had in place for curators to select tags from the tag list was untenable. To remedy this, we updated our tag selection method from a select box, to a select box with a search field so if you knew what you were looking for, you could just search and select your tag, and if you wanted to explore available tags, you still could. Once this change was made for one library, it was possible to deploy the same code base out to other libraries so they could all benefit from this updated experience.

Lastly, as the goal is to have as many institutions involved as possible, costs need to be considered and ideally allow anyone to setup their own instance of the Lentil application. Currently the costs that are involved are an initial setup fee and then a small monthly fee for hosting and image and video backup. The project is also open source, so if the means are available to set the project up in house and save on costs, then that is also a viable option to get involved. All in all though, the overall cost is minimal and should allow most people to try it out with very little risk.

### **Limits on the use of this for libraries**

The amount of staff time required is directly dependent on the number of, and popularity of the selected hashtags used for the harvesting process. Approving images is a simple process, however, choosing the images from the pages of thousands of harvested images can be time consuming depending on how specific the needs of the collection development policy is. It takes time to collate and edit the overall collection to suit, and it may take time to research hashtag trends to keep abreast of any changes or sudden popular trends for images suitable for collection

When someone post an image to Instagram they grant Instagram a licence to use the image (as shown in Appendix 4). Therefore, this type of collecting requires a risk management approach to copyright laws as the person who took the photo owns and retains full copyright of the images posted (see Appendix 4 for Instagram's terms of use).

### **Limits on the technology**

The software that underpins the Instagram archiving project (Lentil) is no longer officially supported. While anyone is still able to edit the code and make any changes they need, the



official developers have declared that they are no longer going to release bug fixes or new features for the project. While the project is fairly stable in its current state, the software that it relies on to run is also slowly no longer being supported, so it's only a matter of time before it becomes a bit of a challenge to run on modern infrastructure.

Like most modern internet based infrastructure, in using Instagram, we're locked in obeying the rules of the platform. While there are literally billions of images on Instagram, we only have easy access to images from the recent past. This means that the sooner we start collecting and archiving images, the better a collection will be.

We're also beholden to the terms of use of Instagram, the company could turn around tomorrow and say that no one is allowed to access their images outside of their mobile application or website. This would mean that the archiving project would come to a halt unless we were able to apply for some special consideration. This is true of any form of information collection though, if the source goes away, we can no longer get the artefacts that we want to archive.

Then there are the hardware limitations of the technology, these can range from simple constraints like where can we get physical storage big enough to hold all of the images we want, to how do we ensure that we're able to make the images available for viewing to the variable size audience that we might have.

### **Working together at a distance**

This project is one which has relied on several communication tools as the participants have never been in the one place, except when online. Email, via NSW public library email lists, was used to call for expressions of interest as well as for other communication. Icelab used Basecamp with each site so that work could be readily tracked. We used Blue Jeans for video conferences to discuss different stages of this work. We have only had a few video conferences but they were vital for connecting people in varied locations across NSW and the ACT. We even used google docs to write this paper. We really have never all been in the the one location. Use of a range of communication tools has been crucial to the effectiveness of this project.

### **Future possibilities**

Collecting contemporary material for local studies collections is a important move for more local studies collections to make. With the volume of content being created each day, it is important for slices of this information to be collected for later research and to reflect a greater diversity of the community. It can also be a key way to show what an area looks like as it changes over time, as well as to record other aspects of daily life in different areas.

Recent content, such as the photographs being collected by Inner West Libraries and Richmond-Upper Clarence Regional Library are showing what their communities look like now. This will keep changing so that beef cattle, snakes, crocheting, chickens, dogs, murals, protests and celebrations will be recorded. These all provide research information about what these communities look like now, and show the value of including social media in a local studies collection.

People share their lives and their stories in many ways. Whose story are we prioritising over another if we choose one format over another? Local history collectors should not limit their collecting based on the longevity of the format, because short-lived technologies tell their

own tale and are constantly being shaped by the communities using them. Future individuals can learn from the context as well as the content.

It is important to note that we are not limited to artefacts from Instagram. There are many more platforms and mediums that people use to communicate about their local surroundings and communities that could be considered in the future. Though photos are easy to quantify and consume (as a viewer), it's also possible to analyse a corpus of Tweets that mention a particular topic or phrase and measure when something becomes popular within a community.

We also want to encourage anyone who is interested in collecting information for local studies to think about how technology can help you as any ideas you have might also help others.

## Appendices

### Appendix 1 Where to find things

Github

North Carolina State University on GitHub <https://github.com/NCSU-Libraries/lentil>

State Library of NSW on Github <https://github.com/slnew/lentil>

Libraries using the software:

James B Hunt Library #myhuntlibrary <https://d.lib.ncsu.edu/myhuntlibrary>

Inner West Libraries <http://images.innerwest.nsw.gov.au/>

Richmond Upper Clarence Regional Library <http://images.richmondvalley.nsw.gov.au/>

### Appendix 2 Expression of interest for second phase of public library involvement

Archiving from Instagram.

Would you like to experiment with collecting material from Instagram for your library?

We are looking for two libraries to test this with us. The aim of this test is to have a straightforward methodology for other public libraries to be able to use.

To participate in this trial, the following points are required:

- Your council or library has an Instagram account
- Your library staff can suggest some hashtags of relevance to the community that they would be interested in collecting - ones in current use, not for a future
- You have the ability to provide library/council header in appropriate format for online use
- You have a willingness to set up external accounts, their setup and login details provided (Heroku, AWS S3, and Postmark)
- You have a willingness to monitor and collect for at least six months – note this will require ongoing staff time of 2 – 4 hours a month.
- There is a dedicated domain or subdomain for the web page to appear on (e.g. <http://community.yourlibraryname.com>) — this could also be a subdomain of your existing library website, or some other shared domain

We already have one test library, but are looking for two additional libraries to sort out some of the issues which emerged.

The James B Hunt Jr Library in North Carolina has been collecting images from Instagram for their archives. You can see some examples here <http://d.lib.ncsu.edu/myhuntlibrary>

You can read about other work they are doing with archiving social media here <https://www.lib.ncsu.edu/social-media-archives-toolkit>

### **Appendix 3 hashtags for Richmond Upper Clarence Regional Library**

#casinonsw, #evanshead, #coraki, #kyogle, #richmondvalleymade, #bundjalung, #casinobeefweek, #richmondvalley, #kyoglecountry, #borderrangesnationalpark, #borderranges, #bundjalungcountry, #lismorefloods, #chscareers, #casinogolfclubmarkets, #stmaryscasino, #bonalbocentral, #kyoglecountrylife, #prdkyoglerealestate, #gatewaytokyogle, #prdnationwidekyogle, #kyoglelifestyle, #kyoglecountryrealestate, #kyogleproperties, #sheepstationcreek, #dobiesbight, #richmondriver, #corakiflood2017, #kyogleflood, #casinobeefweek2017

The RUCRL domain : <http://images.richmondvalley.nsw.gov.au/>

### **Appendix 4 Terms of use statement from Instagram**

“We do not claim ownership of your content, but you grant us a license to use it. Nothing is changing about your rights in your content. We do not claim ownership of your content that you post on or through the Service. Instead, when you share, post, or upload content that is covered by intellectual property rights (like photos or videos) on or in connection with our Service, you hereby grant to us a non-exclusive, royalty-free, transferable, sub-licensable, worldwide license to host, use, distribute, modify, run, copy, publicly perform or display, translate, and create derivative works of your content (consistent with your privacy and application settings). You can end this license anytime by deleting your content or account. However, content will continue to appear if you shared it with others and they have not deleted it. To learn more about how we use information, and how to control or delete your content, review the Data Policy and visit the Instagram Help Center.

Taken from: [https://help.instagram.com/581066165581870?helpref=faq\\_content](https://help.instagram.com/581066165581870?helpref=faq_content) [accessed 7 September 2018].