# How to Manage Privacy in Photos after Publication

Srinivas Madhisetty, Mary-Anne Williams, John Massy-Greene, Luke Franco and Mark El Khoury
*Magic Lab, University of Technology Sydney, Ultimo, Sydney, Australia*

Keywords: Privacy in Photos, Tacit Information, Grounded Theory.

Abstract: Photos and videos once published may stay available for people to view it unless they are deleted by the publisher of the photograph. If the content is downloaded and uploaded by others then they lose all the privacy settings once afforded by the publisher of the photograph or video via social media settings. This means that they could be modified or in some cases misused by others. Photos also contain tacit information, which cannot be completely interpreted at the time of their publication. Sensitive information may be revealed to others as the information is coded as tacit information. Tacit information allows different interpretations and creates difficulty in understanding loss of privacy. Free flow and availability of tacit information embedded in a photograph could have serious privacy problems. Our solution discussed in this paper illuminates the difficulty of managing privacy due the tacit information embedded in a photo. It also provides an offline solution for the photograph such that it cannot be modified or altered and gets automatically deleted over a period of time. By extending the Exif data of a photograph by incorporating an in-built feature of automatic deletion, and the access to an image by scrambling the image via adding a hash value. Only a customized application can unscramble the image therefore making it available. This intends to provide a novel offline solution to manage the availability of the image post publication.

## 1 INTRODUCTION

Currently there have been a few solutions which deal with data immortality. One of the solution has been implemented by Snapchat which is a popular social media application. Its core features is that media such as photos or videos are deleted after 24 hours. (Snap Inc., 2018). This solution requires that the user be online and give control of their data to Snapchat. However while snapchat says it will delete most messages, it also states in various online documentation that Snapchat "can't guarantee that messages and corresponding metadata will be deleted within a specific timeframe" (SMH, 2015).

This research contributes by investigating how to manage privacy in the context of sharing and storing photos.

He et al., (2016) suggests that when images are uploaded to platforms such as Facebook, users are giving up the privacy of the image as they are giving it over to the platform holder.

Social media applications like Facebook, Twitter, WhatsApp and many more applications are becoming popular. The instant sharing of information via photos and videos is making the management of issues which rise out of loss of privacy more difficult.

An example of loss of privacy due to lack of proper photo data management was when security researcher John McAfee was arrested in Guatemala (Finke, 2012).

Identifying sensitive information in a photo or a video is a major problem. For example, what is sensitive to one person may not be sensitive to others. Therefore, rather than making assertions about what is sensitive in a photo this research asked 21 participants why they share content and what are their concerns. The data was analysed using Grounded Theory to determine privacy sensitive information.

This paper discusses the difficulties in managing privacy post publication of a photograph. The idea presented in this paper is at the first stages to develop a comprehensive ontology based on the privacy management, post publication of the content.

Therefore, an offline solution, using Python was developed allowing people to share their images without compromising certain critical aspects privacy. The critical aspects of privacy were determined by asking questions to participants about



Madhisetty, S., Williams, M., Massy-Greene, J., Franco, L. and El Khoury, M.
How to Manage Privacy in Photos after Publication.
DOI: 10.5220/0007614001620168
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 162-168
ISBN: 978-989-758-372-8

their privacy concerns. The collected data from the interviews was analysed using Grounded Theory.

Privacy may be called as a social construct, there are many definitions of privacy. Oxford dictionary defines privacy as "A state in which one is not observed or disturbed by other people".

In layman term's, loss of privacy may be considered when any sharing of information such as photos takes place. This is irrespective of whether the information shared is sensitive or not.

The problem of managing privacy in photographs or videos is such that, for example when a document is shared the context of the document can be easily understood. Such grounding of the context may not be present, when photos are shared. This is because a photos may contain rich semantic and syntactic information coded as tacit knowledge.

This makes it more difficult to manage content as information freely passes through without any checks or balances that are afforded in other means of communication.

By regulating the flow of information in photos and videos, privacy is managed effectively. (Bennett and Raab, 2002) envision a privacy "regime" that integrates privacy policy instruments – including data protection legislation, voluntary fair information codes and privacy-protective information practices - in a global economy which is characterized by regulatory interdependence.

Social networks provide unprecedented opportunity for individuals and organisations to share information. At the same time they present significant challenges to privacy (Chen and Williams 2009).

The main problem is that, after initial publication of the content using social media, its subsequent persistence makes the content not ephemeral. Technology enables the content to be available, such loss of privacy can be attributed to the lack of control about the content published using this relatively new technology.

It may have a significant impact on individual privacy. The ephemeral nature of such information shared it is important to be able to have desirable levels of privacy. For example, when people move on from and into relationships and other major life events, an individual should be able to exercise the right to be left alone. With others able to republish photos and videos using social media the individual's privacy is breached significantly. "People should have the freedom to share whatever information they want, in any medium and any format", the freedom to access all of the information made available to them by others" and "the freedom to build trust and reputation through their identity and connections" (Facebook, 2011).

## 2 THE PRIVACY PROBLEM

The first problem is that people need a reliable method to share photos only with those who are the intended recipient. The second problem is the internet makes images immortal because they can be shared and manipulated. There is a need to introduce mortality or lifespans to images so they don't cause privacy issues in the future.

The final problem is that many people understand that metadata is information about images but don't know how it's used or how to access it. There is a definite need to simply the process of viewing, understanding and controlling the metadata found within photos.

The two main artefacts that were diagnosed from conducting this research were that there was inadequate information about privacy and its consequences after users publishing their content such as a photo or a video.

This research was conducted by asking people why they like to or has shared their photos or videos using social media. By understanding the expectations for publishing content, this research could arrive at a clearer picture about the objective opinion on why participants consider their privacy has been breached.

Questions about the metadata of the photo or video, were asked to understand the tacit properties of the photo or a video.

This did not give the exact contextual properties, but have given a clear indication under which circumstances the photo or video was taken. The information captured was about the shutter speed, ISO, aperture, type of lens being used, etc.

Through this tacit information tagged in a photo it is easy to make inferences about the circumstances the photo was taken at that time.

In order to develop a proof of concept that an application could be built to incorporate the findings which were produced using analysis from Grounded Theory.

The most common form file format for images shared online are JPEGs. JPEGs are common due to the fact that they are a lossy file format. Lossy images are files whose data has been compressed which means that less storage is required to host them. This is an advantage on the internet where data storage costs money.

While JPEG files store the image data they also contain another file format called Exchangeable image file format or "Exif". The Exif file format contains all the metadata about an image and refers to metadata as "tags". These tags include information like owner information, GPS data, times and dates. This information is organized into five groups or "Image File Directories" (IFDs). The IFDs are as follows:

1. 0th - Information that is necessary for the construction of the image.
2. 1th - Information about the construction of the image thumbnail.
3. Exif - This is the main metadata that gives context to the image. It can include a range of extra information such creation date, owner name, camera lens used and even humidity.
4. GPS - Contains location information.
5. Thumbnail - contains the raw bytes of the image thumbnail.

While many of the tags contained in these groups are specified has having a data type such as string or integer some are left as undefined and the implementation can be left up to decision of the programmer. Some tags are also defined as having a range of allowable values.

The maker note and user comment tags which are meant to be used by photographers and manufacturers and are often undocumented.

## 2.1 Research Question

How can metadata in a photo be abstracted such that no further inferences could be made about a when a photograph is shared.

How can privacy be managed using an offline solution post sharing of a photograph?

In order to answer those questions, first it is essential to determine what sensitive information people believe exists in a photo. Second, it is important to associate these findings to assist in developing a framework which will assist the general public to manage their privacy effectively.

It is also equally important to understand the underlying motivations in sharing the photo and to be able to understand its context. These critical features which will allow information in a photo to pass through without affecting its privacy need to be understood and investigated.

This understanding of expectations versus their consequences have given rise to the determinants of privacy. These determinants will manage how the information in a photo will be stored and retrieved. One such determinant to manage privacy was the availability of the photograph, the second determinant was the ability to download the photo and republish the photograph by modifying it.

## 2.2 Motivation and Significance

Using social media where anyone can publish photos and videos of any other individual, mostly well-intended at the time, may result in a privacy concern later.

Once the photo or video is published, it is available for people to see until it is removed by the publisher of that content. During data analysis the key aspect which was determined as a major inhibitor of privacy is the instant availability of information. Such as the availability of a photograph or a video.

Henry et al., (2017) Anastasia Powell said, "her study had uncovered significant levels of image-based sexual abuse. One in 10 adults has had a nude or semi-nude picture of them taken without their permission. The same proportion has had a sexually explicit image of them sent to others without their permission, or had someone threaten to publicly share such an image. "

Revenge porn, has become a big problem in today's society. According to Henry et al., (2017) survey, it is estimated that 1 in 5 Australians have experienced image-based abuse also young people aged 16 to 29 years are also at higher risk of image-based abuse. There is a consensus among the survey that 4 in 5 Australians agree it should be a crime to share sexual or nude images without permission.

An example of image immortality having negative consequences is revenge pornography. Revenge pornography is described as "online release of explicit photographs or videos of an individual without permission for the purpose of humiliation" (Kamal and Newman, 2016).

Kamal and Newman (2016) also state that not only do victims of revenge pornography experience humiliation but they also experience a never-ending struggle to maintain their dignity. There is a need to introduce "mortality" to images and give them a finite lifespan that is only controlled by the owner.

### 2.2.1 Research Methodology

The research design adopted was of two types the first is to gather the evidence via interviews to understand the nature of the problem. The second is to implement the findings from the analysis of the data gathered

into a software program which could be used to manage privacy.

For the first part of this research, twenty one open ended interviews were coded using Grounded Theory. The Straussian approach to develop descriptive accounts in the place of theory development is the approach taken to conduct this study. Strauss and Corbin guidelines in the process of data collection, coding and analysis were used to conduct this research. This approach encourages flexibility to use techniques or steps; it characterizes the situation objectively to obtain a general view from different perspectives as opposed to that of a quantitative study.

Some of the questions were framed in such a way that they repeat themselves. This was done intentionally to get a clearer picture of what the participant is describing to get contextually rich descriptions. Any participant who had a vested interest in technology such as suppliers or producers of web cameras and other electronic devices were omitted as their opinion could be biased, although it is unintentional, but will have a certain impact on the research, i.e. any participant who has a commercial interest in the growth of privacy-inhibiting technologies were also omitted. Open-ended questions are the best approach as they allow the interviewer to obtain tacit information, which was later, contextualized the meaning of the response.

For this research after twenty one interviews sample saturation was reached. All participants interviewed were social media users. They should have uploaded and viewed photos and videos of themselves and others through social media.

The second part was to implement the design via a prototype using Python as a programing language. An offline solution, via software was developed which will allow people to share their images without compromising certain critical aspects privacy.

The developed software will be able to encrypt the image, set up password and a hash value and finally specify a deletion date. When the deletion date is reached the photo gets deleted automatically. The viewer of this photo can only view the content if they have the hash value.

Python was chosen to develop the software because it is suitable for rapid prototyping. The version of python used in the prototype is the latest release 3.6. Below is a list of libraries selected and the reasons as to why they were selected:

- Tkinter - This is a standard library built into python 3 is usable across all major os desktops (Tkinter, 2018).
- Pillow - This is an extension of Tkinter which is well documented and allows for the use of images within a GUI. This library could be extended or removed in later builds after the initial development (Pillow, 2017).
- Piexif - This library is a major part of the project. Testing with this library allowed for Exif data to be cleared as well as added with ease (Piexif, 2015). It should be noted that Piexif is the only library available for Python that manipulates metadata
- Cryptography - This library was used as the basis for allowing image encryption to work. Contains the required tools to generate crypto keys and check their validity. (Cryptography, 2017)

The development and testing has been conducted on a windows device.

### 2.2.2 Key Findings

Timeframe on how long the content is made available was found to be one of the key vectors to manage privacy. For content to be managed effectively a timeline or a timeframe is to be determined for photos and videos before making them available.

Select a timeline for each photo or a video, after which it would automatically disappear from the public space. The photo or video would only be made available for others to see after it had been renewed.

> *Participant p10 "I think a timeline of five to six months is enough for a photo to exist in the public sphere. Managing photos and videos is easier that way."*
>
> *Participant p17* said, "*I think it is a good idea to have a timeline for each photo made public*". After the moment the photo or video is consume there was no reason for it to be available for others to view.

Currently there is no time limit on how many days content is made available to end users.

This research has found that by limiting the number of users who can view the content, or selectively sharing information with a particular group of individuals, is a more effective way to manage privacy when photos and videos are shared.

The interviewers also identified trust and control of information is essential for effective management of privacy. However there was ambiguity in terms of what trust and control actually meant. Control is a simple choice of what information they intend to use to communicate with others, as discussed in (Altman, 1977) view of privacy.

Control meant several things - it was about the type of information or the nature of the information which was sensitive or perceived as sensitive, and also the way the flow of such information that can be managed through various elaborate privacy settings.

## 2.3 Software Solution for Effective Privacy Management Post Publication of a Photograph

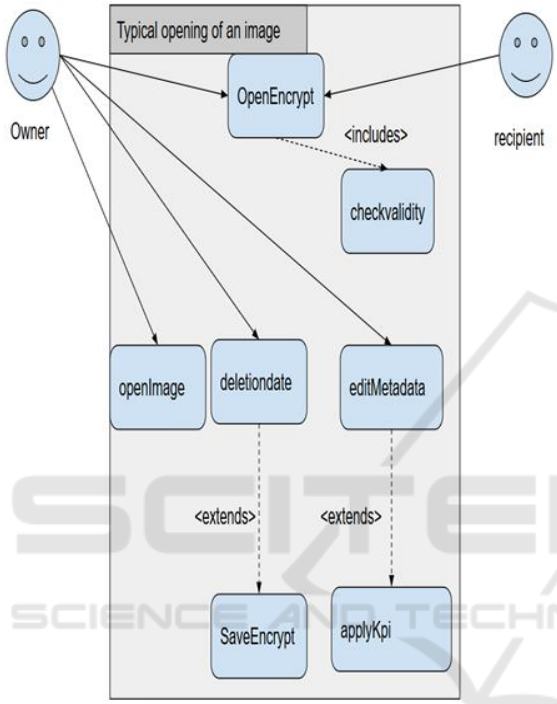


Figure 1: Process to manage privacy in a photo by providing an offline solution.

The proposed solution is a flexible application which can incorporate many more features which are critical to manage privacy. There are other applications which are available that can scramble an image. However, this application was developed to customize the implementation of the Key Performance Indicators which were derived through analysis of interview data using Grounded Theory study.

The owner and the recipient have the hash code to be able to view the image. However, this application not only just checks the hash value, but also checks for a deletion date. The meta-data of the image could be further spoofed or modified to allow other features in managing privacy to be embedded. The other feature that circumvent privacy concerns which may arise due to profiling of metadata embedded in the image is the ability to spoof the metadata. This spoofing of metadata could give rise to inconsistent profile may give more privacy to the end user as it anonymizes the context.

**Data Mortality.**

As data immortality is a problem when addressing image privacy. The software provides a means to implement software mortality.

This is achieved by introducing a date that the image will be deleted through the software. A deletion date will be inserted into the user comment section of the images metadata and checked each time the image is checked. This date is stored in UNIX date format which is single integer number that is easy to compare.

A limitation of this is that the software clock can be manually changed so deletion date may never be reached. In order to overcome this problem there are actually two dates inserted and separated by a double colon, one is deletion date and the second being the date the image was last opened.

This second date is modified each time the image is opened so that it will shorten the life of the image each time that it is open. This strategy allows the program to overcome possible manipulation of the software clocks.

Both these dates are checked when the image is opened. If either date has passed, the image will be considered compromised and then deleted.

**Image Privacy.**

Managing privacy of a photograph while in transit from one user to another requires some form of encryption.

A symmetric encryption scheme will be used as the key can be shared between trusted parties when the image is exchanged. The python library named

Fig 1 Process involved to deliver offline solution for the management of privacy.

Cryptography provides a simple to use encryption scheme that requires a 32-byte key for both encryption and decryption.

This key is generated by the user and then the output is padded with ones in order to meet the 32-byte requirement of the key. The library offers no measures in order to accept a key that is shorter or longer than 32 bytes. The cryptography library creates an object that will encrypt the images data using 128 AES encryption (Cryptography, 2017). This procedure is completed in reverse when the image is opened.

To ensure that the image has not been tampered with a hash is taken before encryption and compared

once the image is opened again to ensure that no alterations have occurred. If there is a difference in the hash then the image is then deemed compromised and deleted from the file system.

**Metadata and Exif Manipulation.**
The ease of manipulating metadata Exif tags. Piexif is a library used to manipulate metadata. Exif tags have specific formats required when being written to. For this reason, the software developed only allows Exif tags that have a finite range of allowable values to be edited. These Exif tags appear in the software as dropdown menus. This design means that the owner of an image will not render the image unusable due to a format error within the Exif data.

To achieve this, a data structure was created that matches all the Exif metadata tags hexadecimal values with the human readable name, as well as the corresponding allowed values that a tag can be. This data structure is contained in tagList.py.

Metadata can also be removed according to 3 levels of KPI's which are called minimum, medium and maximum privacy were afforded.

The deletion of metadata according to KPI's was suggested by the earlier analysis using interview data and Grounded Theory. The functionality has been developed as a proof of concept and not according to specific research.

**Discussion.**
There are hundreds of official tags each with their own data type and allowable values. In addition to this there are countless manufacturer maker note tags that have no official interpretation.

Due to sheer number of official and undocumented metadata available this research allowed to interpret a small set. This was also hampered by the Python library Piexif which is used to extract metadata is not well documented.

To be able manage privacy Exif tags names were examined which were in hexadecimal format. For example the tag BitsPerSample is represented as 0x0102. This means translation was required for every tag and thus the program has a large self-developed translation library. Any tags unable to be translated are referred to as "Undefined".

Another limitation with the Piexif library is that the metadata values can be data structures such as tuples and byte arrays. This is difficult to programmatically render as a string that is understandable to a viewer. Thus, for this version of the program we were required to limit the amount of metadata that a viewer could edit or spoof.

The limitation of being a purely offline solution meant that the software is unable to check online clocks to make sure the deletion date hasn't passed. The software compensates by using the computer's clock to save the date the image was last opened. The next time the image is opened it checks the computer's clock to check if the last time the image was opened was actually in the past and not the future. This was extra checking to make sure the deletion date is honoured.

# 3 CONCLUSIONS

Managing privacy in photo and videos should not be an after thought after the sharing has occurred using social media. As privacy is a loosely defined it is very difficult to for see all the consequences before publishing content. However managing the content effectively will mitigate risks of privacy. The ideal way of managing privacy is to derive a contextual meaning from a photo, and then to draw a conclusion about which photos (or which portions of photos) are appropriate for other users to see, and to manage privacy in the photo or a video before it is shared.

Further research needs to be done to derive an over arching picture about privacy because as technology keeps moving forward, an equal emphasis needs to be given for privacy concerns of individuals. A system that manages user privacy in a digital world is difficult, because privacy is considered by many as a social construct.

This paper presents a software program that gives a proof of concept on how publishers of images can gain control over what to share with others and manage the privacy settings in an offline method.

Users are given control through data mortality and deletion dates as well as passwords and authorization. An offline solution means that there is no need for users to trust that 3rd parties to store their images in an ethical manner. This designed software is flexible for further implementation of features which give more control over the photograph post publication by editing metadata tags. Many such iterations of privacy issues will result in an effective way to retain privacy when photographs are shared on the internet.

# REFERENCES

Altman I 1977, 'Privacy regulation: culturally universal or culturally specific?', Journal of Social Issues 33 (3): 66–84.

Andrews, L. 2012, 'Facebook is using you', The New York Times, 4 February.

Bennett, C. and Raab, C. 2002, Governance of Privacy: Policy Instruments in Global Perspective, Barnes and Noble, London.

Creswell, J.W. 2003, Research Design: Qualitative, Quantitative and mixed methods approaches, Sage Publication, Thousand Oaks, California.

Charmaz, K. 2006, Constructing grounded theory. A practical guide through qualitative analysis, Sage Publication, London.

Chen, S. and Williams, M-A. 2009, ' Privacy in Social networks: A comparative study', PACIS, vol. 4, pp. 81.

Cryptography 2017, *Welcome to pyca/cryptography*, Viewed 23 April 2018, https://cryptography.io/en/latest/.

Facebook 2011, Facebook principles, viewed 17 February 2014, <http://www.facebook.com/principles.php.>.

FC. K. 2010, 'The fundamental limits of privacy for social networks', MIT Technology Review Physics arXiv Blog, viewed 5 May 2010, <http://www.technologyreview.com/view/418819/the-fundamental-limits-of-privacy-for-social>.

Finke, B. 2012, 'John McAfee's location may have been accidentally given away by Vice magazine', *The Telegraph*, 4th December, viewed 28th November 2018, <https://www.telegraph.co.uk/news/worldnews/centralamericaandthecaribbean/guatemala/9720514/John-McAfees-location-may-have-been-accidentally-given-away-by-Vice-magazine.html

Grey, D.E. 2009, 'Doing Research in the Real World', 2nd edn, Sage Publication, London.

He, K., Bidan, C. and Guelvouit, t.L. 2016, 'Privacy Protection for JPEG Content on Image-Sharing Platforms', paper presented to the Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, Vigo, Galicia, Spain.

Henry, N., Powell, A. and Flynn, A. (2017). Not Just 'Revenge Pornography': Australians' Experiences of ImageBased Abuse. A Summary Report. Melbourne: RMIT University.

Stratus, A. and Corbin, J. 1998, Basics of qualitative research, Sage Publication, Thousand Oaks, California.

Smh 2015, *Snapchat-Now-Owns-Your-Photos-Even-After-They-Disappear. Viewed 27 November 2018*, https://www.smh.com.au/technology/snapchat-now-owns-your-photos-even-after-they-disappear-20151102-gkobrl.html

Smh 2015, *'Revenge Porn: Government Urged to Make it Illegal*.' Viewed 10 Nov ember 2018, http://www.smh.com.au/national/government-urgedto-outlaw-revenge-porn-20150926-gjvod5.html

Snap Inc. 2018, *When does Snapchat delete Snaps and Chats?*, viewed 28th November 2018, https://support.snapchat.com/en-US/a/when-are-snaps-chats-deleted

Maykut, P. and Morehouse, R. 1994, Beginning Qualitative Research: A Philosophic and Practical Guide, The Farmer Press, London.

Tkinter, *Python interface to Tcl/Tk*, Viewed 20 May 2018, https://docs.python.org/3.6/library/tkinter.html.

Kamal, M. and Newman, W.J. 2016, 'Revenge Pornography: Mental Health Implications and Related Legislation', Journal of the American Academy of Psychiatry and the Law Online, vol. 44, no. 3, pp. 359-67.

Pillow 2017, *Pillow*, Viewed 19 March 2018, https://pillow.readthedocs.io/en/5.1.x/ .

Piexif 2015, *Welcome to Piexif's documentation!,* Viewed 20 April 2018, http://piexif.readthedocs.io/en/latest/index.html.

Warburton, W.I. 2005, 'What are grounded theories made of? 2005', LASS Faculty Post-Graduate Research Conference University of Southhampton, UK, 6-7 June.