The Integrity of Exif Metadata:

A review of Social Media platform interaction.

Seyed Saad Mazri Asl

Supervisor: Martin Havránek

Czech University of Life Sciences

Prague Faculty of Economics and Management Department of Information Technologies

Abstract

Smartphones provide users access to social media platforms. This allows users to exchange information through messaging, storing, and sharing photographs taken by their smartphones. With smartphone cameras creating images that contain embedded Exif metadata and the use of social media platforms to deliver images to wider audiences, users risk their privacy and safety by sharing the images through disclosure of Exif metadata.

While some users are aware of the metadata functionality, a significant portion of users remain unaware that their camera enabled smartphones have geo- tagging functionality. The geo-tagging functionality embeds the users GPS coordinates into the image file of the photograph they take. The unaware users share additional information when they share their photographs onto social media platforms, specifically geo location information.

This additional information that is disclosed could potentially be used for malicious purposes or evidence in legal proceedings.

This research examined the embedded Exif Metadata in image files are created with the Redmi note 7 smartphone. To verify the integrity of the Exif metadata, the test image files were examined both prior to and after being posted to social media platforms.

CHAPTER 1: INTRODUCTION

This chapter will provide the background information on the main components of this research. It will concentrate on their definitions, as well as the possible consequences brought about by their use and development. The relevance of the study was explored by analyzing the use of smartphones and social media platforms (SMP), the importance of the research becomes relevant through the number of user awareness of service providers' metadata handling practices.

1.1 Background to the study

1.1.1 Metadata

Metadata comes in many forms. The National Information Standards Organization (NISO) describes metadata as "*the information we create, store, and share to describe things, allows us to interact with these things to obtain the knowledge we need."* (NISO, 2017). Metadata defines the content of a file in terms of its condition, consistency, and other technical details including the software and hardware used to create the file, as well as geo-locational data (Clark 2011).

According to NISO (2011), metadata improves the utility of a file by allowing the user to recognize the file's owner, date of creation, and location attributes. When a file's data is considered alone, it can be difficult to interpret and use; but, with metadata, more descriptive knowledge about the file is understood, and the file's utility improves (NISO, 2011).

When you take a picture with a digital camera, it automatically saves Exif data within the photo. Exchangeable Image File Format (Exif) contains the descriptive information stored within that photo. The Exif format is used to embed information such as the time, date, time of capture, camera settings (aperture, shutter speed, focal length, metering model and ISO speed information), copy right information and the location of the capture (CIPA & JEITA, 2012).

1.1.2 Smartphones

According to Loeffler (2021) the first cell phone that was sold commercially was developed towards the end of the twentieth century and was sold in 1983. The cell phone that was made in 1983 was developed for a one single purpose, to make phone calls. As the cellphone was a new technology with a price of 3500$ at that time, it was too expensive for the public to purchase it and as a result, only the wealthiest of people were able to purchase it.

The creation of the first smartphone came to life through IBM in 1992 and released for purchase in 1994. The newly developed Smartphone was not only able to make phone calls but had some additional features that would not only be included in the Simon Personal Communicator (SPC) but every smart phone that followed. The new built in features included, a touch screen, ability to send and receive emails and faxes, a calendar, address book, and a native appointment schedular (Meghan Tocci).

As the accessibility to smartphones increased, so did the demand. Users are no longer bond to desktops as the processing power of a smartphone has become comparable to a desktop computer. Smartphones not only allow us to communicate, but they also help us through navigation, replace the need of physical memory and share experiences through photos (Eran Kinsbruner & Justin Reock, 2020) In 2010, smartphones outsold desktop computers and the use of them led to the development and growth of mobile applications such as Apple store and Google play store (Rakestraw). The mobile application market allowed users to download and share content such as photographs over social networking applications (Janssen, 2010).

With the smartphone adoption, this has encouraged an explosive growth in applications, including third party players that have End User License Agreements (EULAs) that are not ethical. The EULAs are presented in a way that discourages a user from reading the potential risks involved in using the application and how their personal information is processed, thus, the user might agree to potentially share more information than intended through the application (Newitz, 2005).

The most threatening part of privacy that comes from the use of smartphones and applications is the end user location (Humboldt, 2013). Some social media platforms have adopted the practice of stripping away important or all metadata from images uploaded onto social media platforms (IPTC,2016). This resulted in user preference whether to share geo-locational information or not (D. Wassom, 2015).

1.1.3 Accessibility to Resources

The term open source refers to something people can modify and share because its design is publicly accessible (Open Source). The practicality of open source software is that it allows a user(s) to obtain a software with its source code, modify it and then release back to the community. This method of customizing and sharing helps improve the software in the long term as other users can enhance it by adding their knowledge and expertise.

Open source software has had a significant positive impact on cyber security. Open source operating software such as Kali Linux and Parrot Linux are offensive security distributions that come with a range of pre-installed tools. As these tools are available to the public to benefit from them, these tools are also available to users with malicious intent. Users with malicious intent can use tools such as metagoofil and Exif Pilot to harvest metadata from images whether it may be from SMPs or other alternatives to use them for a variety of purposes, such as selling them to the highest bidder on the black market.

There is an abundant amount of information and step by step tutorials that can be found on the internet on how to use these tools. Examples of such accessibility to resources are YouTube channel by David Bombal and an educational cyber security platform called Tryhackme. The educational channel and platform provide step by step tutorials and hands on experience on how to utilize various tools to perform activities such as data harvesting, Denial of service (Dos), network scanning/compromising and create/modify malware.

1.1.4 Location privacy

In an era where data is considered as valuable, if not more valuable than a nations currency, user location plays a significant role. Social media platforms utilize user geo-location to provide targeted ads, recommend friends nearby, deliver local news, track user routine/places they visit and, in some cases, help law enforcement track criminals or assist in finding a missing individual.

The possibilities of utilizing the geo location extracted from the metadata are broad. Whether a user's geo location is extracted from the image(s) they upload or IP address or nearby cell phone tower(s), users end up sharing more information when they share a post than they are aware of. Majority of social media platform users are young or technically unaware, this, combined with the significant amount of personal information they share allows malicious individuals or organizations to exploit them. According to the International Labour Organization, the human trafficking business kidnapped over 20 million people, of whom are mostly women, and generated an estimate of $150 billion in profits in 2014. There are over 500,000 online predators active each day. Therefore, location privacy matters.

1.2 Problem Statement

Exif metadata is embedded in images and it provides information about the file, specifically the geo location information. This metadata can be modified and lose its integrity once uploaded onto a social networking site. The lose of this data would be inadmissible if retrieved from a social networking site for legal proceedings.

This research examined at how GPS functionality in smartphones can be used to embed geo-locational information in the camera's output. It's essential to map out how much information the location-based functionality embeds in files despite the user's preferences. It's also vital to verify whether Exif location-based metadata is relayed and stored in the SMPS, regardless of the users phone settings. There is a significant number of users worldwide that are compromising their privacy without knowing it.

1.2.1 Research Aims and Objectives

This research will focus on the integrity of Exif metadata that is embedded in a photo. It will determine if any modifications occurred when the files interact with a social media platform. If the results of the finding suggest that the files have been modified, the integrity of the data would be inadmissible in courts. It is also focused on raising awareness about metadata practices in hopes to encourage people to make more informed decisions when it comes to sharing information on these platforms.

The research process involved assessing to what extend do smartphones embed metadata into images that are produced from smartphone cameras, as well as asses whether the Geo-location of these images are exported and stored in social media platforms, despite the user's preference.

1.2.2 Research Questions

The aim of this study will be to raise awareness about metadata, the interaction between a user uploading an image onto a social media platform and to answer the research questions below.

1. Is the integrity of Exif metadata in the images maintained after the images are uploaded onto the social media platforms or are they stripped away?
    - Can variables such as account privacy settings have any impacts on the Exif metadata?

2. How can the Exif metadata be affected if the original image file is uploaded again with modified Exif metadata?
    - Can the authenticity of the Exif metadata be proven?

3. What are the practices social media platforms upholding?

CHAPTER 2: METHODOLOGY

2.1 Introduction

This chapter focuses on how to implement the experiment and collect the necessary data to address the study questions. The study's research questions and objectives require multiple experiments to determine how the variables impact the integrity of Exif metadata. Different Social media platforms use different metadata handling methods; therefore, it is crucial to determine whether the metadata has been modified with because of the varying metadata handling strategies.

2.2 Selected Methodology

A quasi-experimental approach was used to address the research questions proposed earlier. This procedure will isolate and monitor all applicable conditions that affect the events under evaluation. The study's related conditions are image files with defined Exif metadata and SMP accounts with differing privacy settings. The observations are recorded during the study and when the defined conditions are modified, the observations are then compared to the initial Exif metadata in the pre- test files. This procedure will be repeated to address the difference of both the default and maximum privacy settings accounts.

2.3 Proposed Research Design

The Exif metadata will be controlled and divided into two. The first image will contain the unmodified Exif metadata and the second will image will contain the modified data. Both images contained the same Exif metadata prior to being modified.

CHAPTER 3: LITERATURE REVIEW

3.1 Introduction

The following chapter reviews literature on the integrity of Exif metadata files that interacted with SMPs. It also reviews literature on metadata and geo-tagging.

The aim of this chapter is to reveal how Exif metadata is handled across different social media platforms, the factors that disclose metadata, the policies social media platforms implement.

3.2 Metadata

Metadata is descriptive information embedded in files that proves to be useful for querying, archiving and additional information purposes (Zhang & Gourley, 2009). The availability of embedded metadata allows a user to find out when and where an image was taken, as well as camera lens properties and even the copyright information of the photographer (Fletcher, 2009). Copyright records, image descriptions, capturing system details, and timestamps are all stored in metadata formats.

Since metadata is stored in its own file inside a directory, multiple metadata formats can be combined in a single file. As each of the metadata formats stores specific fields of data, this means that metadata can be overlapping. Though Exif is the most widely used metadata format in digital photography, other metadata formats exist such as International Press Telecommunications council (IPTC) and Extensible Metadata Platform (XMP) (Gumerov,2012).

3.3 Geo-Tagging

Geo-tagging refers to the process of adding locational metadata to a media output such as an image file (Finjanmobile, 2017). This information is stored as metadata in the Exif format and allows users to determine where the image was taken by using the longitude and latitude coordinates that are embedded in the Exif metadata.

3.4 Previous Studies

3.4.1 Embedded Metadata Manifesto

Members of the IPTC Photo Metadata Working Group examined how social media sites handle metadata embedded in photos posted to their sites. The aim of this three-year analysis, from 2013, 2016 and 2019 was to see whether embedded metadata is stripped off when photos are uploaded to social media sites (IPTC, 2019).

An image file with embedded metadata, including Exif metadata, served as the place to start. After that, the test image was submitted to a number of SMPs. A series of tests were performed after the image was uploaded. The first test was to see if all of the image file's embedded metadata fields were visible on the SMPS's web interface. If the metadata was visible in the SMP web browser, it was verified for any inconsistencies.

The second test was to check if the metadata displayed the 4C columns, Creator, Caption, Copyright Notice or Creditline. This stage was irrelevant to my research as my research focuses on geo-location fields.

The third step was to download the image by right clicking on the image and selecting save image as. If this was possible, then the image would be compared to the original image that was uploaded to the SMPs.

The fourth step process was to repeat the download of the image using the SMPs user interface. If the feature was present, the downloaded image would then be compared against the original image for embedded metadata.

The findings of the previous studies (IPTC, 2013, 2016, 2019) showed that the metadata was still present in the downloaded image files for several of the social media platforms that were checked. Also, in certain instances, the metadata was displayed on the user interface of the social media sites. Following the report, the group concluded that there was a need to establish more effective metadata management strategies.

CHAPTER 4: Practical

4.1 Pre-test Observations

4.1.1 Image 1 – Unmodified

Below is the recorded observation for the first controlled category (the unmodified image file).

Filename: Test Unmodified.jpg

File MD5: 7f5c52c74f1b663c11196b7eb51ce884

GPS values

Latitude: 50.062075

Longitude: 14.445265

4.1.2 Image 2 – Modified

Below is the recorded observation for the second controlled category (the modified JPEG image file).

Filename: Test Modified.jpg

File MD5: d62bed1534c3bcdb2e5aa67bb0a0167a

GPS values

Latitude: 52.062075

Longitude: 16.445265

```
Brightness Value                : 2.18
Exif Image Width                : 4000
Exposure Mode                   : Auto
Aperture Value                  : 1.8
Components Configuration        : Y, Cb, Cr, -
Color Space                     : sRGB
Scene Type                      : Directly photographed
Shutter Speed Value             : 1/50
Exif Version                    : 0220
Flashpix Version                : 0100
Resolution Unit                 : inches
GPS Latitude Ref                : North
GPS Longitude Ref               : East
GPS Altitude Ref                : Above Sea Level
GPS Time Stamp                  : 11:40:04
GPS Processing Method           : CELLID
GPS Date Stamp                  : 2021:03:12
X Resolution                    : 72
Y Resolution                    : 72
Make                            : Xiaomi
Thumbnail Offset                : 3413
Thumbnail Length                : 7065
Compression                     : JPEG (old-style)
Image Width                     : 4000
Image Height                    : 3000
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
Aperture                        : 1.8
Image Size                      : 4000x3000
Megapixels                      : 12.0
Shutter Speed                   : 1/50
Create Date                     : 2021:03:12 12:40:12.466316
Date/Time Original              : 2021:03:12 12:40:12.466316
Modify Date                     : 2021:03:12 12:40:12.466316
Thumbnail Image                 : (Binary data 7065 bytes, use -b option to extract)
GPS Altitude                    : 255.1 m Above Sea Level
GPS Date/Time                   : 2021:03:12 11:40:04Z
GPS Latitude                    : 50 deg 3' 43.47" N
GPS Longitude                   : 14 deg 26' 42.95" E
Focal Length                    : 4.7 mm
GPS Position                    : 50 deg 3' 43.47" N, 14 deg 26' 42.95" E
Light Value                     : 7.0
```

```
Flashpix Version                : 0100
Color Space                     : sRGB
Exif Image Width                : 4000
Exif Image Height               : 3000
Interoperability Version        : 0100
Sensing Method                  : Unknown (0)
Scene Type                      : Directly photographed
Exposure Mode                   : Auto
White Balance                   : Auto
Focal Length In 35mm Format     : 0 mm
Scene Capture Type              : Standard
GPS Latitude Ref                : North
GPS Longitude Ref               : East
GPS Altitude Ref                : Above Sea Level
GPS Time Stamp                  : 11:40:04
GPS Processing Method           : CELLID
GPS Date Stamp                  : 2021:03:12
Padding                         : (Binary data 2060 bytes, use -b option to extract)
Compression                     : JPEG (old-style)
Thumbnail Offset                : 7578
Thumbnail Length                : 4918
About                           : uuid:faf5bdd5-ba3d-11da-ad31-d33d75182f1b
Creator Tool                    : Windows Photo Editor 10.0.10011.16384
Image Width                     : 3000
Image Height                    : 4000
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
Aperture                        : 1.8
Image Size                      : 3000x4000
Megapixels                      : 12.0
Shutter Speed                   : 1/50
Create Date                     : 2021:03:12 12:40:12.466316
Date/Time Original              : 2021:03:12 12:40:12.466316
Modify Date                     : 2021:03:12 17:25:46.466316
Thumbnail Image                 : (Binary data 4918 bytes, use -b option to extract)
GPS Altitude                    : 255.1 m Above Sea Level
GPS Date/Time                   : 2021:03:12 11:40:04Z
GPS Latitude                    : 52 deg 3' 43.47" N
GPS Longitude                   : 16 deg 26' 42.95" E
Focal Length                    : 4.7 mm
GPS Position                    : 52 deg 3' 43.47" N, 16 deg 26' 42.95" E
Light Value                     : 7.0
```

4.1 Post-test Observations

This part consists of the data gathered from the images throughout the experiment procedure.

The images that were uploaded to the test accounts were accessed using an incognito window of Google Chrome.

Flickr

Direct upload to Flickr – Test Unmodified		
Test Unmodified	Filename	51031541838_765261cf69_o.jpg
	File MD5	7f5c52c74f1b663c11196b7eb51ce884
Observations	The images uploaded to Flickr presented with a few options to obtain the images. If users right clicked and saved the image, they would be presented with a HTML file and a folder containing a few sizes of the images; these images contained no Exif metadata. However, by clicking the download button on the Flickr website I was able to download the image sets that still contained the Exif metadata The downloaded images contained all the correct Exif metadata although they had been renamed for database/archiving purposes. The MD5 hash values for both the file and the GPS coordinates did match the originals indicating there was no modification at all and the integrity of the metadata is maintained.	
Test Modified	Filename	51031541873_7e14d5ef9b_o.jpg
	File MD5	d62bed1534c3bcdb2e5aa67bb0a0167a
Observations	The images uploaded to Flickr presented with a few options to obtain the images.	

	If users right clicked and saved the image, they would be presented with a HTML file and a folder containing a few sizes of the images; these images contained no Exif metadata. However, by clicking the download button on the Flickr website I was able to download the image sets that still contained the Exif metadata The downloaded images contained all the correct Exif metadata although they had been renamed for database/archiving purposes. The MD5 hash values for both the file and the GPS sub-IFD did match the originals indicating there was no modification at all and the integrity of the metadata is maintained.

Facebook

Direct upload to Facebook – Test Unmodified		
Test Unmodified	File name	160095766_2521171658176809_3128145784919899450_o.jpg
	File MD5	b4e98d52343d95b3a4224708a50bcbbc
Observations	The images directly uploaded to Facebook were renamed and the metadata was stripped from the files. On both the default privacy settings and the maximum privacy settings I was not able to access the original images nor the metadata stored in the images.	

Instagram

Direct upload to Test Unmodified – Default Security Settings		
Test Unmodified	Filename	159586388_189637912615445_5942972436223451784_n.jpg
	File MD5	e7cee5589034fa5e5c643f529d145852
Observations	Instagram presents very limited options for users.	

	The image hosted on Instagram contains no information, is a low quality and resolution snapshot of the originally uploaded image. The privacy settings only changed the option to have a public or private profile which meant users would request permission to follow and see the images. There are no options to download or save the image and even inspecting the element for an image source only provides the snapshot generated by Instagram. Due to the nature of this site and its primary functionality only available via a smartphone app, there was no method of saving the image via smartphone app.

Tumblr

Direct upload to Tumblr – Test Unmodified		
Test Unmodified	Filename	tumblr_d5fedf8ed3cc47e67bbdfc372d2b1428_7956b00d_2048.jpg
	File MD5	c5cc7b5624bdd228953800e6c95ea7dd
Observations	The images directly uploaded to Tumblr were renamed and the metadata was stripped from the files. On both the default privacy settings and the maximum privacy settings I was not able to access the original images, nor the metadata stored in the images.	
Direct upload to Tumblr – Test Modified		
Test Modified	Filename	tumblr_b4e921a98b42d1e7c5352d7cb2e128dc_e909d740_2048.jpg
	File MD5	8809ad2c2e03620dc156ab1300cbbea4
Observations	The images directly uploaded to Tumblr were renamed and the metadata was stripped from the files. On both the default privacy settings and the maximum privacy settings I was not able to access the original images nor the metadata stored in the images.	

LinkedIn

Direct upload to LinkedIn – Test Unmodified		
Test Unmodified	Filename	1615642976681.jpg
	File MD5	85724e676351d07e10c4731aa45373df
Observations	The images uploaded to LinkedIn appear to be a new JPEG image file that contains only the image data. For both the original unmodified and modified images that were uploaded contained the same image data but differed in Exif metadata. The images obtained from LinkedIn have different file names as well as different hash values.	
Direct upload to LinkedIn – Test Modified		
Test Modified	Filename	1615642973372.jpg
	File MD5	9b73c25a23f224516c30ca097c78784e
Observations	The images uploaded to LinkedIn appear to be a new JPEG image file that contains only the image data. For both the original unmodified and modified images that were uploaded contained the same image data but differed in Exif metadata. The images obtained from LinkedIn have different file names as well as different hash values.	

Pinterest

Direct upload to Pinterest – Test Unmodified		
Test Unmodified	Filename	download.jpg
	File MD5	d8aeca59517c7122156e2b91fa91003b
Observations	The images uploaded directly to Pinterest allowed for the images to be downloaded by right clicking and saving the image. For both the original unmodified and modified images that were uploaded contained the same image data but differed in Exif metadata.	

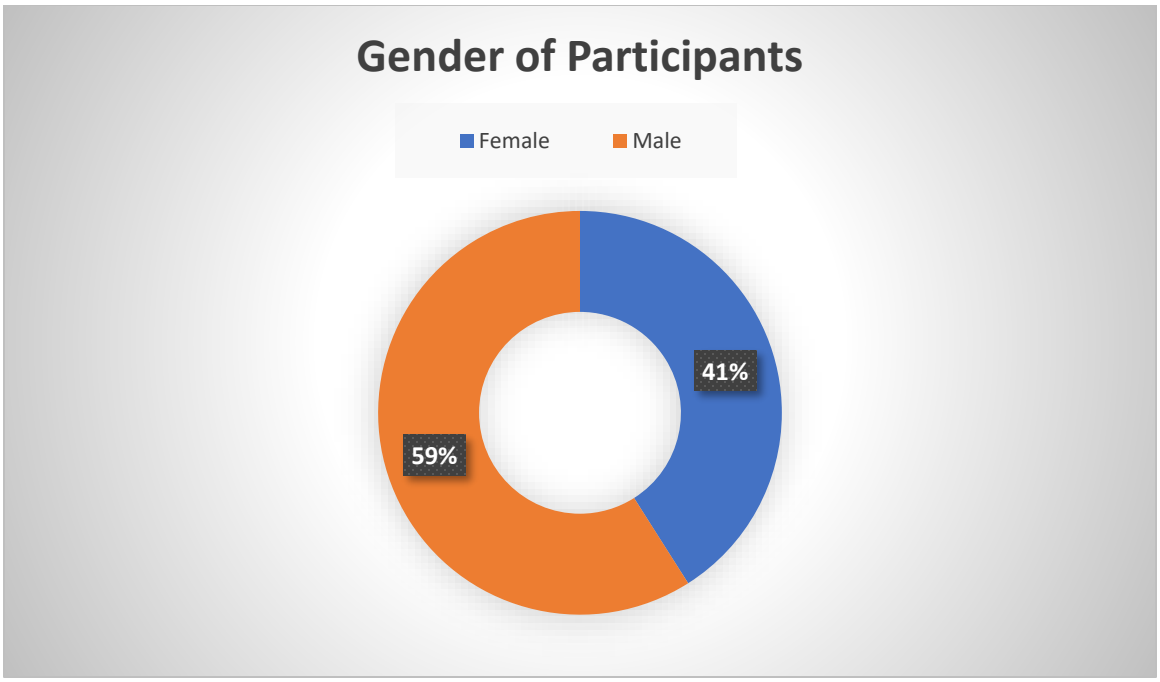
	The images obtained from Pinterest have different file names as well as different hash values.	
Direct upload to Pinterest – Test Modified		
Test Modified	Filename	download.jpg
	File MD5	b90b71d5350856c22a0def93e2c2d7b9
Observations	The images uploaded directly to Pinterest allowed for the images to be downloaded by right clicking and saving the image. For both the original unmodified and modified images that were uploaded contained the same image data but differed in Exif metadata. The images obtained from Pinterest have different file names as well as different hash values.	

Twitter

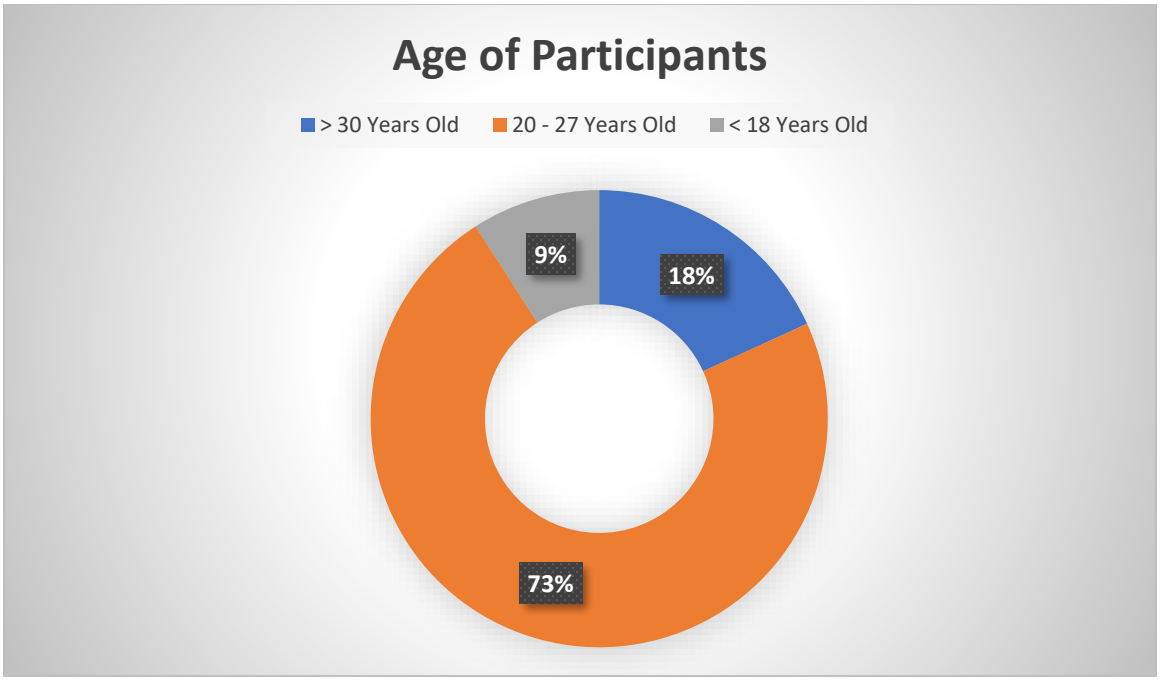
Direct upload to Twitter – Test Unmodified		
Test Unmodified	Filename	EwXBwZWW8Awfn5K.jpg
	File MD5	251e96d5cc02169e483b06662c358a82
Observations	The images directly uploaded to Twitter had the metadata wiped and only retained the image data. While the default account settings allow anyone to view the post, the location is disabled by default. If the uploaded image contains Exif metadata, Twitter can display an accurate location based on the coordinates provided, if the location feature is enabled. However, downloading the images would only provide the images that have the metadata stripped.	
Direct upload to Twitter – Test Modified		
Test Modified	Filename	EwXBwZOW8AAXmGC.jpg
	File MD5	a44a58bde40028e8bacf8d527d899866
Observations	The images directly uploaded to Twitter had the metadata wiped and only retained the image data.	

Exposure

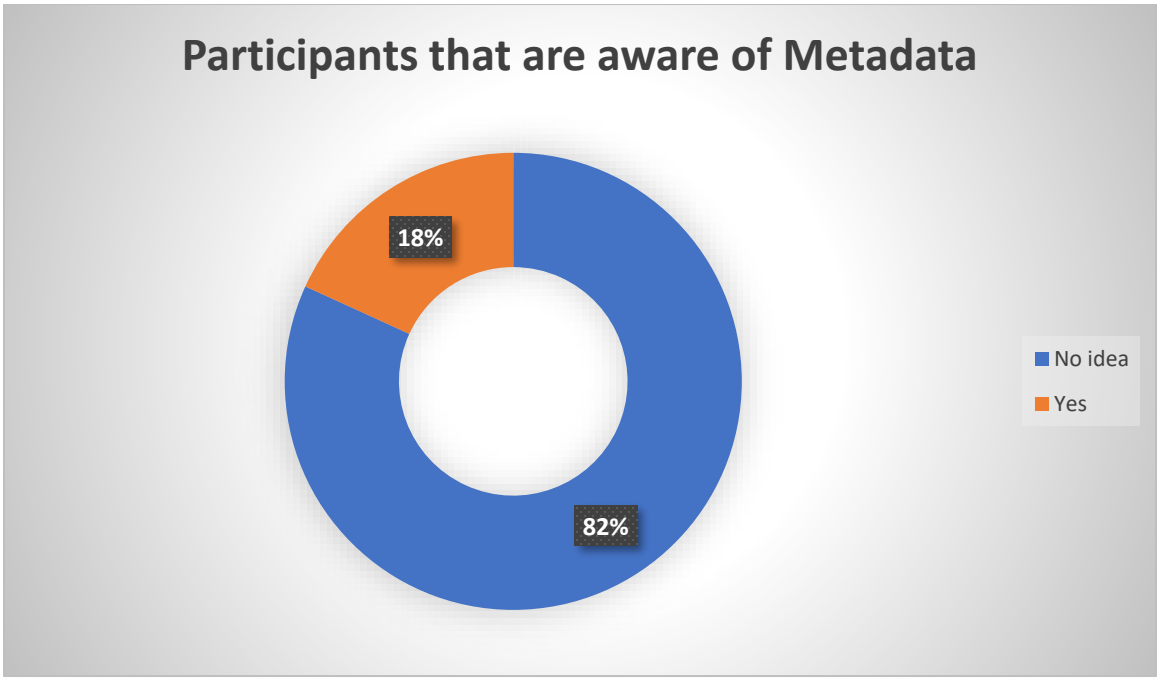
Direct upload to Exposure – Test Unmodified		
Test Unmodified	Filename	original.jpg
	File MD5	7a62fdccc12108413427bd3c1a44a9d1
Observations	The images uploaded directly to Exposure allowed for the images to be downloaded by right clicking and saving the image. For both the original unmodified and modified images that were uploaded contained the same image data but differed in Exif metadata. The images obtained from Exposure have different file names as well as different hash values.	
Direct upload to Exposure – Test Modified		
Test Modified	Filename	original.jpg
	File MD5	aecfe08e376fc4591f2495ee1e18f615



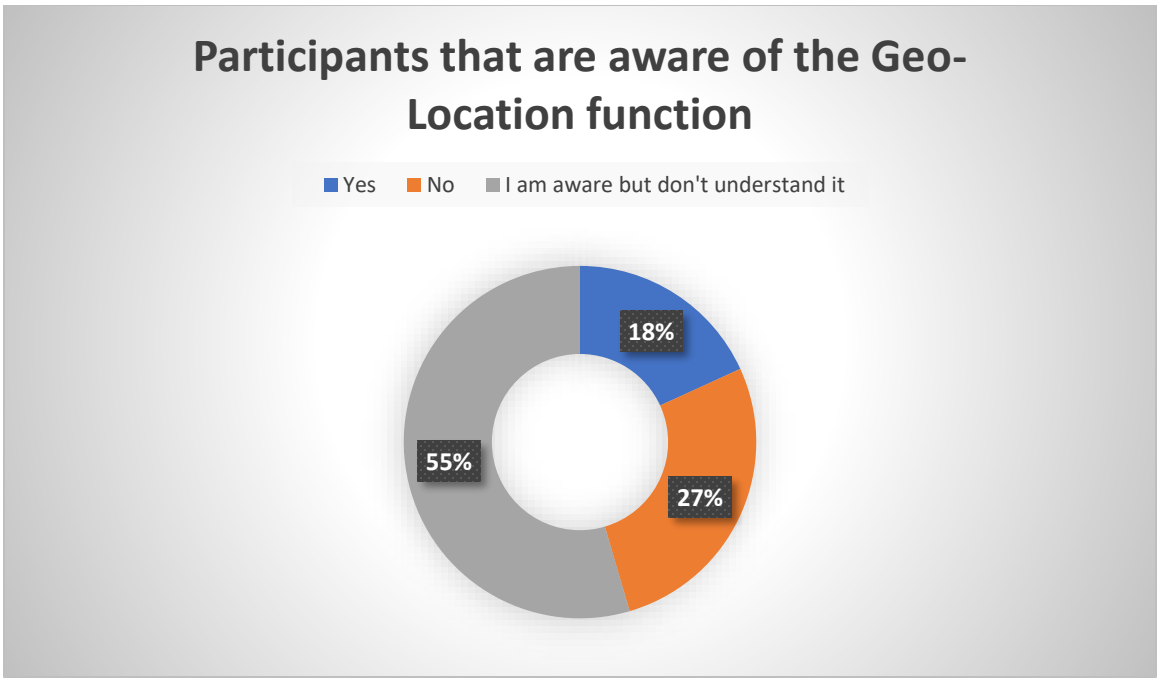
The gender of voluntarily participated in this research. Female gender is highlighted in blue (Right). Male gender is highlighted in brown (Left).



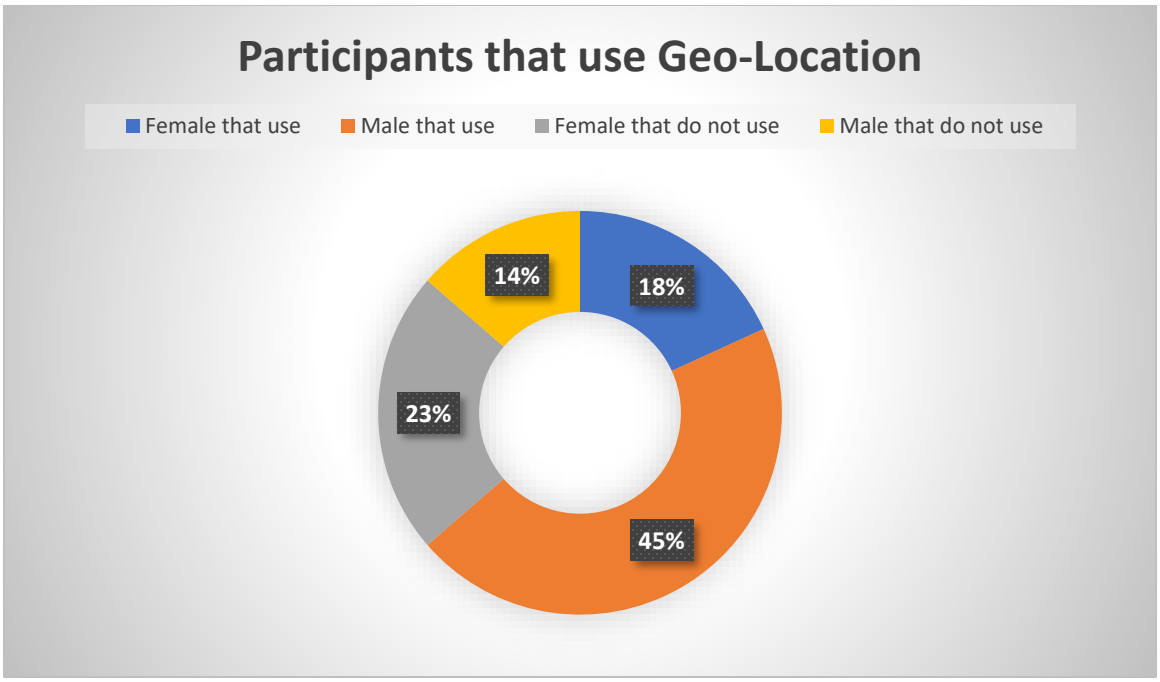
The age of the participants: Majority of the participants ages 73% range between 20 to 27 years old (Brown). 18% of the participants that are older than 30 years old (Blue) and the minority consisting of ages younger than 18 years old (Grey).



Participants that don’t understand or have never heard about metadata: 82% have never heard or don’t understand metadata (Blue) and only 18% of the participants are aware about the topic metadata (Blue).



Participants that are aware of the geo-location function that embeds the GPS coordinates in the photos they generate: Only 18% of the participants responded with a Yes (Blue), 55% (Grey) of the participants had some idea but did not understand the technology and how the function was able to obtain these coordinates, leaving us with the rest of the participants 27% (Brown) that were not aware at all.



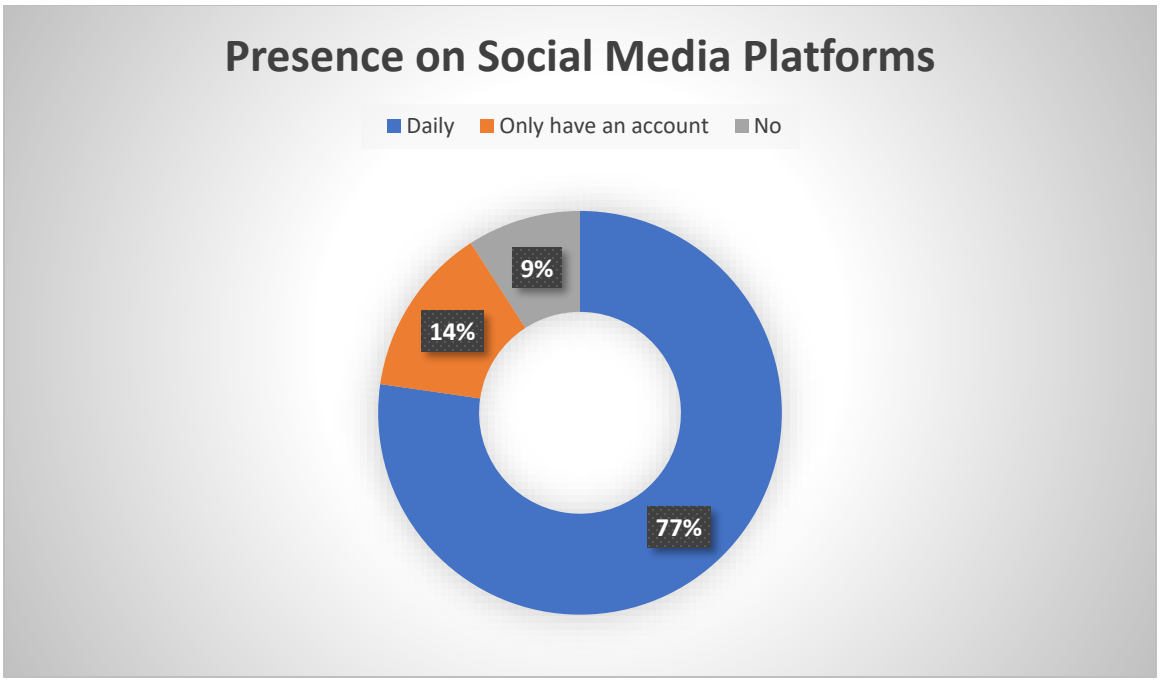
Participants that share their geo-location when uploading media on to social media platforms:

Females that do choose to share their geo-location: 18% (Blue)

Males the do choose to share their geo-location: 45% (Brown)

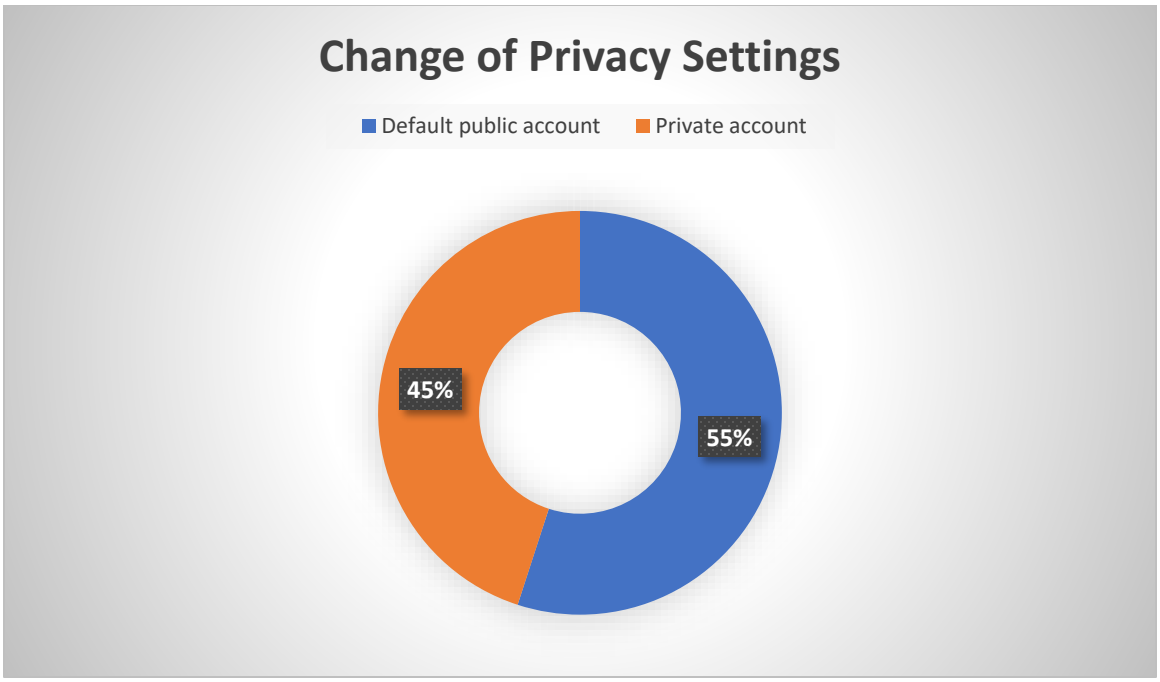
Females that do not share their geo-location: 23% (Grey)

Males that do not share their geo-location: 14% (Grey)



Participant's presence on social media platforms:

Majority of the participants 77% (Blue) use social media platforms daily, while 14% (Brown) just have an account and 9% do not have any presence on any social media platform.



Participants that changed their privacy setting from the default public account to private account:

The results in this section show that 55% (Blue) of the participants did not make any changes to their security settings, while 45% (Brown) did change their security settings over time.

CHAPTER 5: Results and DISCUSSION

5.1 Introduction

The following chapter will discuss the findings of the experiments carried out in this research. It will discuss any impacts that occurred to the Exif metadata embedded within the image files that were posted on the SMPs. This chapter will also discuss the results of the survey that was conducted.

5.2 Metadata Handling Practices

The majority of the SMPs demonstrated consistent handling practices. Almost all the SMPs modified the images by stripping the metadata except for the image data itself. The social media platforms that were tested provided different levels of customization of account privacy. Some of the sites allowed the user to download the original image without stripping away the metadata.

One of the sites provided an opportunity to download the images without stripping away the data was Flickr. Flickr provided the option to download the images with a range of sizes that included the GPS coordinates.

Some of the social media platforms that were experimented on, such as Facebook, Instagram, Twitter, LinkedIn did provide a geo-location tag option to share with other users when uploading the images. This function was optional, as the user could choose to either share the geo-location of where the image was taken from or not. The result of this means that social media platforms do store the metadata of files that are uploaded.

The social media platforms that were tested provided varying levels of customization for the account privacy settings. Instagram, Pinterest, Tumblr only provided a public or private setting, while other sites offered more security/privacy settings. Facebook and Flickr allowed privacy settings on specific images and albums.

Although Flickr did allow privacy settings on specific images and albums, the images were available to download in a range of sizes including the original image for default accounts. The original image that could be downloaded from the same page included the Exif metadata that was embedded in the image. The only difference observed when disabling sharing Exif information was that the Exif information would not be displayed on the page but nonetheless, users could still download the image and would be able to view the Exif information within the image itself.

The table below outlines the different handling procedures discovered in this research.

Social media platform	Strength/Weakness		Hash Values
Facebook	Strength: • Images are renamed. • No metadata could be found stored in the images.	Weakness: • The Metadata is stored prior to being stripped from the images and are available through the account data file.	• No Match.

Flickr	Strength: • Right click and save option is unavailable. • Images were renamed.	Weakness: • Download image(s) button available. • Downloaded images contained the Exif Metadata.	• Identical Exif information
Tumblr	Strength: • Renaming and changing the hash values.	Weakness: • N/A	• No Match
Instagram	Strength: • Download or save option are unavailable. • No traces of information could be found within the images. • Users have limited options.	Weakness: • Low quality when a snapshot of the original is taken.	• No Match
LinkedIn	Strength: • Renaming of images. • Metadata is stripped away.	Weakness: • Does not differentiate between the images, both images contained the same MD5 file hash value.	• No Match
Pinterest	Strength: • Possible to limit the audience for image boards. • Renaming of images. • Changing hash values.	Weakness: • Exact metadata retrievable.	• No Match

Twitter	Strength: • Metadata is stripped away from the images.	Weakness: • Default account allows anyone to view tweets and posts.	• No Match
Exposure	Strength: • Renaming and changing the hash values.	Weakness: • Right click and save image available. • Both sets of images contained the same image data but differed in Exif metadata.	• No Match.

5.3 Survey

The survey consisted of 22 participants, 13 male and 9 female. The results obtained from the survey are as following:

Majority of the participants could be categorized as young individuals. The first noticeable observation was that most of the participants 82% were not aware about metadata and only 18% were, although the number of participants is small, it is still visible that social media platforms do not educate their users enough.

Moving on to the geo-location function that is incorporated in phones and generates the GPS coordinates in photos, undeniably, due to the fact there are GPS applications such as Google Maps, 55% of the participants did confirm they are aware that their phones contain a geo-location function, however, they do not understand how the technology works.

As mentioned earlier, most of the participants are young individuals who are excited to share their moments in the social media scene. Social media platforms encourage users to share the location of the media they want to upload by automatically filling in the location of the media during the process of uploading. We can observe that most of the participants do share their geo-location upon uploading their media. However, another interesting observation could be made, the number of females that share the geo-location compared to males is drastically lower.

As technology advanced/advances, it's difficult not to participate with the rest of society in utilizing technology. 77% of the participants use social media platforms daily as it is part of their life. While 14% just have an account and 9% do not have any presence on social media. This proves that social media platforms have been adopted rapidly in a short amount of time since their existence.

The final question of the survey focused on participants that changed their account security settings. 45% of the participants changed their security settings, while 55% remained using the default public account. Based on these results, people are not concerned about sharing their privacy even though they are aware they are sharing their geo-location with the rest of the world. The participants that understand the privacy settings still disclose their geo-location regardless of if their profile was private or not.

CHAPTER 6: CONCLUSION

6.1 Thesis overview

Chapter 1 of this research provides a background on key components of the study which elaborate on metadata, smartphones, accessibility to resource and the importance of location privacy. It introduces the factors that disclose metadata such as the awareness of geo-tagging and the default procedures. It identifies the reason why social media adoption is increasing by linking the adoption of smartphones. It also highlights the accessibility resources that could be used to harvest metadata and the consequences of having poor privacy.

Chapter 2 focused on outlining the required methodology that suited the research. The selected methodology allowed the testing of account privacy settings by using two image sets, one unmodified image, and one modified. The first phase involved generating the image sets, verifying the metadata, specifically the GPS coordinates and the MD5 hash value that were generated. The second phase of the procedure was to upload them onto the two types of accounts that were created, one with maximum security and the other with the default public account.

Chapter 3 focused on reviewing previous studies that assisted the author in conducting this research. The studies revealed that there was a need to establish more effective metadata management strategies as there was a lack of user awareness and a large amount of photos that were posted contained Exif metadata.

Chapter 4 consists of two practical parts. The first part is the verification of the generated test images and the detailed procedures of the pre-test/post-test observations. The second part of the practical experiment focused on the survey that consisted of 22 participants in which the results revealed showed that most users are not aware about metadata and the consequences that could be brought forward due to lack of privacy.

Chapter 5 discussed the results of the experiments in which the first experiment revealed the metadata handling practices social media platforms uphold. The strength and weakness between the platforms on how they handle metadata and the results of the second experiment (survey) in which in the results revealed that most users are not aware about metadata, geo-location functionality that is part of their smartphones and the lack of care about their privacy due to the fact it is not encouraged to learn how they can be more careful with their data.

6.2 Summary

The discovered results of the experiments carried out in this research provide an understanding of how social media platforms handle metadata and the lack of user awareness when sharing sensitive data. Whether the user altered the media before uploading or the SMP stripping away the metadata, the results obtained after retrieving the images prove that the integrity of the metadata is not intact. This result makes the integrity of the metadata unreliable in legal proceedings. Furthermore, the research revealed the time consumption and extend a user must take to keep their privacy. In comparison between the social media platforms, the SMP that stood out in the case of how they handle metadata was Instagram. Instagram stripped away the metadata and did not allow permission to right click and save image, nor did it have an option to download the media. This practice should be a benchmark for other SMP to adopt.

The research findings highlight the need for a standard protocol to be adopted by all social media platforms when handling metadata, the protocol should strip away the metadata to avoid being obtained but store the original media with the Exif metadata for legal proccedings.

REFERENCES

https://www.ilo.org/global/topics/forced-labour/lang--en/index.htm

https://opensource.com/resources/what-open-source

https://medium.com/kidguard-education-and-publishing/social-media-kidnapping-9441cb946d08

Fisher, D., Dorner, L., and Wagner, D. Location Privacy

(2014). Metadata for Digital Asset Management. Retrieved from Adobe: https://docs.adobe.com/docs/en/aem/6-1/administer/content/assets/metadata.html

S., Eckles, D., Good, N., King, S., Naaman, M., & Nair, R. (2007). *Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing*. http://www.rahulnair.net/files/chi07-ahern-mediaprivacy.pdf

(2005). *Metadata Awareness Among Information ProfessionalsA Case Study On National Library Of Malaysia* Retrieved from http://eprints.uitm.edu.my/4461/1/MOHAMMAD_AZHAN_BIN_ABDUL_AZIZ_05_24.pdf

(2008). *Privacy Perceptions of Photo Sharing in Facebook*. http://cups.cs.cmu.edu/soups/2008/posters/besmer.pdf

(2012). Authenticating Social Media Evidence. *New York Law Journal*, 248(65). http://www.paulweiss.com/media/1211973/4oct12tt.pdf

(2014). *Metadata and the Law: What Your Smartphone Really Says About You.* Retrieved from Real Clear Technology: http://www.realcleartechnology.com/articles/2014/03/03/metadata_and_the_law_what_your_smartphone_really_says_about_you_1007.html

(2013). *Social Media Investigation for Law Enforcement.* Oxford: Elsevier Science.