Digital Photos, Embedded Metadata and Personal Privacy

Rutgers University has made this article freely available. Please share how this access benefits you. Your story matters. [https://rucore.libraries.rutgers.edu/rutgers-lib/54997/story/]


Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

SOAR is a service of RUcore, the Rutgers University Community Repository
RUcore is developed and maintained by Rutgers University Libraries
Introduction

When individuals organize and curate their own content, the ethical ramifications are generally the last thing considered. This is particularly true considering that most older adults doing such curation have spent most of their time creating content in a pre-digital age. In the past, analog artifacts were limited by their media in how far and wide they could be shared. Until very recently, photographs, sound and moving image recordings used to be primarily physical, and creating them required dedicated, discrete equipment (a camera or recorder, and film or tape) as well as time, money and labor (film developing and tape editing) to yield a finished product. The photographs and recordings themselves were limited in their tangible nature as well: for the most part, sharing these memories meant showing or passing it around to a limited group of people in the same physical space, or having additional copies printed. As a result, the forethought and planning required often limited the spontaneity and volume of these artifacts, creating a limiting factor that kept the number of artifacts we created to an inherently low number, relative to what is achievable on more accessible, digital media. Meanwhile, the relatively limited sharing potential of analog formats made it possible for embarrassing, undesirable or compromising photos or recordings to be more easily hidden or restricted in access, for the most part.

The landscape of documenting our personal histories – photography in particular – changed with the rise of digital creation tools. Their ease of use and additional features meant that not only could people create more photographs and recordings at massive volumes, but the information that these items contain are richer in detail, and provide ever more granular data about our daily lives. How we share these memories has changed as well, relying on third party social networks to instantly distribute our memories – and the rich data they contain. Locations, dates, times, even the type of camera or smartphone you use and the software installed on it, are widely distributed with everything we share. This provides opportunities not only for our family,
friends and acquaintances to know more about our lives, but can provide useful and profitable information to the third parties with which we entrust our data.

These issues of data and privacy give rise to a whole new set of ethical concerns that must be considered as we archive and curate our own data, but are often overlooked or ignored by most. On the other hand, outright blocking of this data collection and boycotting the features and services of these social networks deprives the user of important and useful tools that can be legitimately used to organize and make sense of an otherwise impossible to manage collection. Making personal preservationists aware of the issues, and educating them on both the benefits and pitfalls of data sharing, can help them make informed decisions on the ethical matters surrounding their content, and what level of disclosure is prudent for them.

The primary ethical dilemma: Data sharing and privacy

The new capabilities afforded by high quality digital cameras, smartphones capable of high-speed data, and social networks have no doubt allowed for a transformational shift in how people interact with each other. Users of these tools now have the capability of engaging in and maintaining intimate social relationships with people anywhere, at any time, that are purposeful and deliberate. However, many of these same users that enjoy the beneficial aspects of the technology are concerned with the detrimental aspects. In fact, social network providers such as Facebook are keenly aware that there at least a segment of their users who will actively modify their behaviors – including what they access and how – out of concerns for their privacy. Such behaviors include preferring not to install a social media app on their phone, out of an awareness of that app’s ability to collect data, threatening their sense of privacy.¹

Most creators of digital content are primarily concerned with the substance and message of the work they’re creating – the image they are trying to capture; the content of the document they’re typing, drawing or editing. What may or may not be know is that a wealth of additional, supporting metadata is being captured alongside the actual content of the work, and this metadata


is being embedded into the digital file being created. It is this data which can store personal information, and in some cases this data may inadvertently be leaked by the tools we use to share and organize our personal digital archives.

Automatic data embedding: what your digital camera is learning about you

Fig. 1: A digital photograph captured using a smartphone, and the technical metadata (Time, date, location, smartphone make and model) embedded within the photo.

The integration of high-quality cameras into smartphones has resulted in a fusion of these technologies, into something that collects a great deal of data about the images captured, and by definition, the user of each device. In addition to a camera, many modern smartphones also come equipped with a GPS receiver; a clock that is constantly kept to the correct time using cellular phone network signals; a compass which can determine what direction the camera was facing; an accelerometer which can determine movement and speed; and an altimeter showing elevation.³ By default, the information from each of these sensors gets collected automatically, and can be embedded in each photo and video taken by a smartphone, using a standardized metadata schema for digital photographs known as the Exchangeable Image File Format (EXIF). The result is digital photograph or video which not only witnesses the scene before it, but also describes, with incredibly precise detail, the exact location where the photographer was standing and in what

direction, the exact date and time the photograph was taken, and even the exact make and model of the smartphone itself.

Social media, precise metadata, and privacy implications

Smartphone apps, particularly those used by social media or by retailers, can also access location and usage data. Such apps then apply that information to provide the user with a context-sensitive display, or tailor search results and content creation options. “Check-ins” also broadcast to your social group where you’ve been, and potentially who you were with at the time:

Content organizing tools and social media platforms also rely a great deal on user interaction and crowd-sourcing, to gather more nuanced information. Most users of Facebook are familiar with “tagging,” or digitally associating themselves and other users along with places, actions and even moods, and applying these tags to created content such as a written status post, or a photograph or video. Potentially even more definitive is facial and image pattern recognition based on such tagging. Social media platforms such as Facebook routinely analyze photographs uploaded by their users, and suggest automatically tagging individuals whose faces appear in each image, based on tags applied to previous photos. Similar functionality exists for photo organizing and sharing services like Google Photos, which even takes the capability a step farther, and can recognize common objects like “trees” or “cars” within photographs and index the photos under these contextual tags for later searching.

---


Positive Aspects of Metadata

To be fair, many of the features discussed in this chapter were designed to be positive and useful to the end user. Although the initial reaction upon first learning of these privacy and ethical concerns might be to disable all forms of metadata collection, some careful consideration should be given for the benefits that these same points of data can afford a personal archivist in their quest to organize and make sense of their digital collections. With the average mobile phone user taking as many as 150 pictures per month, such metadata, invasive as it may seem, is often a vital and powerful tool to keep personal media collections from remaining a disorganized mess. Such metadata encoding, for instance, can easily permit an archivist to recall all photos taken in a specific location or date range; facial recognition algorithms can be used to rapidly weed out photos taken of a specific person, or as a memory jogger when events call for it. Such actions can be done far more rapidly than having an archivist painstakingly review hundreds - or even thousands - of photos to manually organize them. Similarly, a person wanting to curate their social media postings will likely find it convenient to search for content they’ve created while they were visiting specific places, or while being with certain friends and acquaintances. Arguably, these are the well-intentioned uses of these features that make them worth integrating.

Negative Aspects and unintended uses of digital content

Much of the consternation about creating and sharing digital content is that it may share too much, and provide a level of information about our whereabouts, habits, affiliations and preferences that are beyond what a person could be comfortable with. This can be particularly true when one tags their friends as appearing in certain photos, and being in certain places,

without their prior knowledge or consent, thus making it more difficult for individuals to directly control what information is shared about them, and to whom that information is shared.

Much of the concern surrounding the use of these technologies, and the potential unwanted exposure of private information that can result, is fueled by very detailed demonstrations of the ability to retrace an individual’s steps, and carefully track their travels, activities and routines, based solely on the data collected by their personal devices and social media content.¹ Users have a tendency to make use of these convenience-adding features where location and other metadata are used to better organize their collections, only to react with surprise and shock when they fully grasp the scope of the information collected, and how that data can be used to paint a picture of that person in an unsettlingly detailed fashion.

Rights, Ownership, and Commoditization of content

Adding to the ethical conundrum are the potential commercial aspects of this data, and how it can be monetarily exploited. Social Media platforms like Facebook, Twitter and Instagram, as well as online content organizing services like Flickr and Google Drive/Photos, ostensibly provide much of their services at no cost to end users. However, these services are run by for-profit companies, whose investors expect them to eventually collect an income. Naturally, the users of a particular service – and the unique content they provide – become a readily exploitable resource for monetization.

Most of these platforms provide a very clear explanation of the types of data they use and the methods with which they use it. However, the explanations aren’t often in places where users are most likely to see them. Most users of Facebook, for example, are provided with a small link on sign-up which give them the opportunity to view Facebook’s data policy before agreeing to it. However, based on self-reported user data, fewer than 16% of social media users claim to always read such privacy policies, and at least 12% never do.¹¹ Meanwhile, up to half of Americans may not understand what a privacy policy is or does.¹²

---


Even so, if a user does decide to peruse Facebook’s data use policy, how their data is utilized is very starkly laid out:

Depending on which Services you use, we collect different kinds of information from or about you.

- **Things you do and information you provide.** We collect the content and other information you provide when you use our Services, including when you sign up for an account, create or share, and message or communicate with others. This can include information in or about the content you provide, such as the location of a photo or the date a file was created. We also collect information about how you use our Services, such as the types of content you view or engage with or the frequency and duration of your activities.

- **Things others do and information they provide.** We also collect content and information that other people provide when they use our Services, including information about you, such as when they share a photo of you, send a message to you, or upload, sync or import your contact information.

- **Your networks and connections.** We collect information about the people and groups you are connected to and how you interact with them, such as the people you communicate with the most or the groups you like to share with. We also collect contact information you provide if you upload, sync or import this information (such as an address book) from a device.  

Unfortunately, leaving the platform and deactivating your account does not make this data usage go away, as Facebook does reserve the right to retain user data even after they’ve stopped using it.  

To be sure, Facebook is not alone in this pervasive data mining. Twitter and Google have privacy policies in place which spell out similar practices in data aggregation and usage.

**Knowing and mitigating the concerns**

Admittedly, avoiding unwanted data sharing is an extremely difficult task, despite it presenting significant potential ethical quandaries. Most of the hardware and software we use for creating born-digital content is configured to “bake in” a significant amount of metadata that tells a great deal of the history of the object, including what some may deem sensitive, personal information. These features are by large created with good intentions, providing powerful and useful tools for organizing, sorting, searching and curating the content. However, disclosing this data may have unwanted consequences, such as disclosing places people visited and at what time, or who may have authored sensitive data when they may have wanted to be anonymous. Some of these unwanted disclosures could be mildly embarrassing, or could be seen as a

---


nuisance when individuals are barraged with targeted advertising. Or, such disclosures could expose individuals to serious scrutiny, such as in cases where activists are making a political statement or performing activities not condoned by their government.

When performing digital curation activities, or arranging the born digital content on our own personal collections or those of others, all of these concerns will need to be considered, along with the potential consequences of sharing this data. The potential risk and negative aspects of those consequences will need to be balanced against the positive benefits provided by location and other personal metadata, and the platforms which make use of them. Once users are made aware or educate themselves of the ethical and privacy concerns that abound, they must make decisions which strike a balance comfortable for them, between security, privacy and convenience, by balancing which services and pieces of data they are comfortable with sharing against the potential results of undesired data sharing."

Some editing methods and alternative platforms do exist to mask data, or avoid some of the privacy and ethical pitfalls befalling the more commonly-used services. These include:

- Editing or wiping embedded metadata on digital files before they are shared. Software packages such as exiftool (multiplatform, http://www.sno.phy.queensu.ca/~phil/exiftool/) and AnalogExif (multiplatform, https://sourceforge.net/projects/analogexif/) permit users to view and edit the embedded information contained in various file types. Doing this will of course change or obliterate information that could be legitimately useful when organizing the content, or trying to search for it later.

- Eschewing common, commercial, and online-based tools for organizing digital resources in favor of open-source platforms. Digital Asset Management tools such as Razuna (http://www.razuna.org) and Stuff Organizer (http://stufforganizer.sourceforge.net) can be used in place of the more common media platforms to curate content. However, setting up this software does require some knowledge in servers and infrastructure, and the convenience advantages associated with using the wider commercial platforms are lost.

---

Conclusion

Education and consent play a key role in bringing awareness to technology users of these issues, and reducing or possibly even alleviating concerns about what personal data they may accidentally be “leaking” to the internet at large. When collecting and curating digital content, gaining knowledge of the type of metadata encoded in each resource, including what information these items may have about other people, will permit the curator to get a grasp of what potential impacts disclosing that information, including who beyond the creator of the work may be affected due to identity tagging or the use facial recognition tools. Making certain that affected individuals are aware of the tagging, and perhaps ensuring that these individuals find that to be acceptable, could help to avoid unwanted surprises when the collection is shared with others later.

As information professionals, we should encourage patrons to:

- **Familiarize themselves with technical metadata.** Digital photographers and content creators understandably focus on their artistic expression, and less so on the technical nature of their output. There are tools built into the software and devices they use which can show them what is embedded in their digital photos, from the location data to anything personal they may want to filter out.

- **Not panic.** When a patron is just realizing that location and other identifying metadata is being shared, a common response is to scrub their photos and digital collections of this information, and to take steps to disable recording it.

- **Understand situations where this technical metadata is useful.** While knowledge and information control is certainly necessary, there are situations where this data is actually very useful for curating one’s personal collection, and where desired features and functionality are lost in the rush to anonymize. Location, time, and other embedded metadata is incredibly helpful in organizing our content, jogging our memories, and rapidly retrieving photos from parts of our growing digital collections. As our smartphones grow bigger memories and our personal photograph collections grow into the thousands, finding “that one photo” can grow increasingly difficult without having metadata as a finding aid.
• **Make “sanitized” copies of photos and content.** As information professionals, we should make ourselves familiar with software tools used to read as well as remove metadata from digital content. Patrons should be educated that these tools can be used to keep control of their personal data, and to make sure that photos that don’t contain sensitive information are shared on social media, when appropriate. And, rather than removing that data from their original files, they should adopt practices similar to those used in the digital preservation community: keep archival copies intact, and make separate, derivative “presentation” copies that contain all of the edits and modified information they wish to share.

We have a unique opportunity to better educate digital content creators about how embedded metadata can affect them, in both constructive ways, and in ways that may cause them privacy or ethical concerns. Holding seminars or classes on digital photography and creating content, as well as having one-on-one sessions with patrons, can go a long way in making aware of these issues, and having fruitful conversations with patrons about digital photos, content and privacy matters.

**Bibliography**


