

# Embedded Metadata and the Digital Lifecycle of Images: Methodological Challenges

Nathalie Casemajor

Université du Québec en Outaouais  
283, boulevard A.-Taché C.P. 1250,  
succ. Hull J8X 3X7 Gatineau Canada  
1 819 595-3900 #2005  
nathalie.casemajorloustau@uqo.ca

## ABSTRACT

This paper analyses the specificities of metadata embedded in photographic images. It investigates how embedded metadata can help studying the usage patterns and conditions of circulation of images on digital networks.

## Keywords

Embedded metadata; image; circulation; usage traces.

## 1. INTRODUCTION

Embedded metadata is a set of descriptive, technical and administrative information that travels within digital files. By recording the marks created by human activity and image processing devices, embedded metadata allow interoperability between platforms, operations of classification and copyright management. For social science researchers interested in the usage patterns of digital content, such collections of traces are wonderful corpora that document the lifecycle of pictures, providing clues on the context of creation, use and circulation of images. They can help reconstruct the material biographies of digital objects.

In the field of media studies, the literature is lacking a methodological consideration of image embedded metadata and their potential for social science. This paper analyses the specificities of metadata embedded in photographic images: how can they be used to analyse the circulation of digital images on the Web? What are the limits and methodological challenges of this type of data? The aim of this paper is to contribute to the field of metadata studies by highlighting the specificities of embedded metadata as an object of inquiry, and by exploring its relationship to the technical, social and economic conditions of digital images circulating online.

Copyright is held by the author/owner(s).  
WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.  
ACM 978-1-4503-4144-8/16/04.  
<http://dx.doi.org/10.1145/2872518.2891071>

## 2. METADATA AS SOURCE OF INFORMATION

Broadly speaking, metadata is “data about data”: a layer of information that describes the structure and attributes of electronic resources. A type of digital record, metadata can take many forms: “system metadata, file system metadata, application metadata, document metadata, email metadata, business metadata, geographical metadata and many more” [1].

This paper focuses on document metadata informing on the “context of how, when, who and in what form the [document] was created or modified or accessed” [1]. Examples of document metadata include the date of creation of a digital photograph, the author’s name, or the last time the file was modified. For Bohm and Rakow, document metadata contains six major types of information: representation of media type, content description, content classification, document composition, document history and document location [2].

The main types of image metadata standards that can be useful to study the social biographies of images are *Exif* (generated by cameras), *XMP* (generated by applications) and *IPTC* (generated manually by users). *ICC* (dedicated to color profile management) is more reduced in scope, but it can hold information on the trajectories of images: for example, Facebook embeds its own color profile on all images that transit through its platform.

Kirschenbaum, Lee, Woods and Chassanoff suggest that digital forensics methods, and particularly metadata analysis, can advance the study of an artefact’s provenance and chain of custody. Forensic tools can extract metadata associated with images, but also traces of user logins or user accounts. For example, the IP address of a user who consulted a resource online, or the number of times a document was downloaded from a webpage.

Embedded metadata is a valuable source for information to analyse the spatial and temporal flow of digital images. Archiving usage traces, they bear evidence of the modes of production, commoditisation, dissemination, consumption and appropriation of digital content. Between trace and operator of movement, they can be analysed as “circulatory devices” [3]. Studying the pattern of dissemination of a digital image on the Web in a comprehensive manner would mean to capture and analyse the multiple set of social and technical interactions that compose image usage, including viewing, displaying and reproducing files.

### 3. CAVEATS OF METADATA ANALYSIS

There are numerous caveats in metadata analysis, and it is important to remember that inaccuracy and information loss is common in images collected online. These errors and lacks can lead to potential misinterpretation, and it is difficult to evaluate the level of risk in each image. Chances of error are especially high with time stamps and dates.

First, the initial camera settings may be wrong, and second, when the device travels, it does not record any indication of its current local timezone, rendering difficult to interpret its correspondence with Coordinated Universal Time (UTC) [4]. Third, as the Sedona Principles point out, there is a “real danger that information recorded by the computer as application metadata may be inaccurate”, because simply moving a file “from one location to another may change creation or modification dates found in the metadata” [5].

Fourth, when images are uploaded to web platforms, much of their metadata is removed, especially in the smaller versions and thumbnail formats generated by web applications such as Facebook, Twitter, Instagram, Wordpress and Flickr. Privacy may be one of the reasons why this information is removed, but the main purpose is to lighten files and facilitate the storage and display of images online.

Scarce but easily collectible online, and readily analysed with automated tools, embedded metadata is a complex research material, composed of technical and social traces of interaction. The major caveats of these corpora are time stamp inaccuracy and lack of integrity (frequent loss in data schemes). These impediments need to be carefully assessed in methodologies seeking to analyse the digital lifecycle of images. A thorough knowledge of the conditions of creation of images, as well as complementary content analysis and context analysis, can help reducing the risk of misinterpretation. But such a thick description is only applicable to small corpora, leaving open the methodological challenge in the case of big data projects.

### 4. CONCLUSION

Image embedded metadata analysis can contribute to better understand the social life of digital things, but its scope is limited to a fraction of these interactions (file origin, modification, and traces of transit through web platforms). Its methodological framework also needs to be further developed in the field of media studies. Some of its critical issues concern the shortcomings of metadata corpora as well as the ethical questions related to metadata collection and exploitation.

Researchers also need to critically reflect on the implications of using unobtrusive methods otherwise used for forensic and surveillance purposes. Embedded metadata analysis is enmeshed in the microphysics of power and control that shapes digital techniques [6] [7]. As such, it contributes to trouble the relationship between informed consent and flows of personal information available online.

### 5. REFERENCES

- [1] Raghavan, S. 2013. Digital forensic research: current state of the art. *CSI Transactions on ICT*, 1, 1, 91-114. DOI: <http://dx.doi.org/10.1007/s40012-012-0008-7>
- [2] Böhm, K. and Rakow, T. C. 1994. Metadata for multimedia documents. *ACM Sigmod Record*, 23, 4, 21-26. DOI: <http://dx.doi.org/10.1145/190627.190635>
- [3] Scott, J. 1990. *A Matter of Record: Documentary Sources in Social Research*. Polity Press, Cambridge, England.
- [4] Boyd, C. and Forster, P. 2004. Time and date issues in forensic computing—a case study. *Digital Investigation*, 1, 1, 18-23. DOI: <http://dx.doi.org/10.1016/j.diin.2004.01.002>
- [5] Sedona Working Group on Electronic Document Retention & Production. 2007. The Sedona Principles (Second Edition): Addressing Electronic Document Production. The Sedona Conference. Available at: [http://sos.mt.gov/Records/committees/erim\\_resources/A%20-%20Sedona%20Principles%20Second%20Edition.pdf](http://sos.mt.gov/Records/committees/erim_resources/A%20-%20Sedona%20Principles%20Second%20Edition.pdf) (Accessed 4 December 2015).
- [6] Foucault, M. 1975. *Discipline and Punish*. Vintage, New York.
- [7] Graham, S. and Wood, D. 2003. Digitizing surveillance: categorization, space, inequality. *Critical Social Policy*, 23, 2, 227-248. DOI: <http://dx.doi.org/10.1177/0261018303023002006>