

12-1-2012

Atomism and Automation

Molly Shaffer Van Houweling

Follow this and additional works at: <http://scholarship.law.berkeley.edu/btlj>

Recommended Citation

Molly Shaffer Van Houweling, *Atomism and Automation*, 27 BERKELEY TECH. L.J. (2012).
Available at: <http://scholarship.law.berkeley.edu/btlj/vol27/iss3/8>

Link to publisher version (DOI)

<http://dx.doi.org/https://doi.org/10.15779/Z38CD8P>

This Article is brought to you for free and open access by the Law Journals and Related Materials at Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley Technology Law Journal by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.

ATOMISM AND AUTOMATION

Molly Shaffer Van Houweling[†]

TABLE OF CONTENTS

I.	INTRODUCTION	1472
II.	COPYRIGHT ATOMISM IN THE DIGITAL AGE	1473
III.	PHOTOGRAPHS AS ATOMISTIC COPYRIGHTED WORKS	1476
IV.	STRATEGIES FOR MITIGATING THE COSTS OF ATOMISTIC OWNERSHIP	1480
	A. REINVIGORATING REGISTRATION.....	1480
	B. DIGITIZING DECENTRALIZED NOTICE.....	1481
V.	ANXIETY ABOUT AUTOMATION	1487
	A. INTEROPERABILITY.....	1487
	1. <i>Technical Interoperability</i>	1487
	2. <i>Legal Interoperability</i>	1490
	B. PERSISTENCE.....	1492
	C. USER COMPREHENSION.....	1496
	D. ACCURACY.....	1497
VI.	CONCLUSION	1500

© 2012 Molly Shaffer Van Houweling. Originally published in the Berkeley Technology Law Journal. Licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900 Mountain View, CA 94041.

[†] Professor of Law, University of California, Berkeley. For helpful discussion and comments on previous drafts, thanks to Robert Barr, Robert Glushko, Robin Kuntz, Mike Linksvayer, Bennett Marks, Deirdre Mulligan, Marianna Murch, Pamela Samuelson, Jason Schultz, Paul Schwartz, Jennifer Urban, and participants in the Berkeley Digital Copyright Project's Symposium on Orphan Works & Mass Digitization: Obstacles and Opportunities. I serve as a Faculty Director of the Berkeley Center for Law and Technology, which supported this research in conjunction with a gift from Nokia (which is a member of the Metadata Working Group, discussed below). I also serve as a member of the Board of Directors of Creative Commons, which is also discussed below. The views I express here—and of course any errors—are my own.

I. INTRODUCTION

Imagine: A budding amateur photojournalist captures the aftermath of a devastating storm using his mobile phone. He then uses the phone to enter a few terms that describe the photo (“hurricane,” “flood”), checks a box to indicate that he wants the photo to be displayed to the public, and checks another to indicate that he is the copyright owner and grants to the public a license to reuse the photo so long as it is attributed to him. He then hits a button to upload the photo wirelessly to the web server of his favorite photo sharing site, where it is immediately available for anyone to view, download, and redistribute via their own favorite photo sharing sites, social networking applications, and news outlets covering the disaster.

Wherever the photo goes on this digital journey, it carries information about its contents, its origin, and its ownership—including the date, time, and location where it was taken; the descriptive terms the photographer added; and his copyright ownership, license, and contact information. When the photo appears online—whether on the photo-sharing site where he originally posted it, or on another photo-sharing site, a social networking platform, a news service, or a peer-to-peer network—this information is communicated to anyone who might want to use the photo yet again. This “metadata” is digitally embedded in the photograph in a way that also makes it possible to find the photo easily by entering the descriptive terms or the photographer’s name into a search engine. The photographer uses this functionality to track the many places online where his photo has been used—with credit to him and his reputation. He receives email inquiries from news outlets who are interested in seeing more of his work.

Technology-savvy photographers know that much of what we have just imagined is not imaginary. Uploading a time-, date-, and place-marked photo from a smartphone to the web while adding descriptive terms that will be displayed along with the photo is becoming commonplace. It is also increasingly commonplace for online photos to be accompanied by copyright licenses that specify how they might be reused. And all of this information can be embedded in digital photo files in a way that can be recognized by subsequent applications and users. But although individual elements of this vision exist, and some have been combined in ways that offer a glimpse of the entire picture, digital photographs do not yet easily and reliably travel on their digital journeys accompanied by information that accurately describes and identifies them and their copyright status. Instead, despite the remarkable potential of digital capture technology and the Internet, digital images can be as mysterious as long-forgotten photos found boxed up in the attic, missing their context and human connections. They become part of our increasingly and frustratingly “atomistic” copyright environment, in which

copyrights are numerous, widely-distributed among often unidentified owners, and fragmented into small and idiosyncratic parts that complicate—or even foreclose—negotiations over reuse of copyrighted works.

In this Article I use digital photographs as a case study of copyright atomism and how automated systems for tagging and tracing might help to alleviate atomism's costs. Part II introduces the copyright atomism concept and explains how atomism has increased in the digital era. Part III focuses on digital photography as an example of extremely atomistic copyright. Part IV describes how embedded metadata might help to alleviate the costs of atomism in this context, but Part V identifies challenges to the metadata solution. Part VI concludes with thoughts about the role of law at the intersection of atomism and automation.

II. COPYRIGHT ATOMISM IN THE DIGITAL AGE¹

Digital technology enables individuals to create and communicate in ways that were previously possible only for well-funded corporate publishers. This makes copyright law newly-relevant to individual amateurs (as opposed to corporate publishers and professional artists and authors) in two different ways. First, individuals are more likely than in the past to reuse copyrighted materials in ways that implicate copyright law—by posting copyrighted photographs on Facebook or other digital platforms without the owner's authorization, for example. Second, individuals are more likely than in the past to use these same digital platforms to publicly distribute *their own* copyrighted creative works. As many observers have noted, most provisions of copyright law in the United States and elsewhere were not designed with these digital-era developments in mind.² Instead, copyright law is too complicated, arcane, and counterintuitive to be well understood by its new targets and beneficiaries.

To be sure, copyright law has changed in recent decades, and some of those changes—specifically, the removal of any formal prerequisites for

1. This Part draws heavily on Molly Shaffer Van Houweling, *Author Autonomy and Atomism in Copyright Law*, 96 VA. L. REV. 549 (2010).

2. Jessica Litman has been an especially insightful observer of this misfit. See Jessica Litman, *Revising Copyright Law for the Information Age*, 75 OR. L. REV. 19, 22–24 (1996); see also Molly Shaffer Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535, 1539 (2005); Pamela Samuelson et al., *The Copyright Principles Project: Directions for Reform*, 25 BERKELEY TECH. L. J. 1175, 1177 (2010) (“Copyright has, of course, always touched and enriched untold ordinary people, as well as specialists, by fostering the provision of entertainment, education, and other information goods and services, but until recently copyright law was relatively invisible to the general public. Amateurs as well as professional artists and authors are now encountering copyright issues on a regular basis.”).

copyright protection—were motivated in part by a desire to make copyright more accessible to individual authors and artists.³ When copyrights are automatically bestowed on individual creators who retain those rights (as opposed to assigning them to intermediaries), the creators have an opportunity to exercise a degree of authorial autonomy that befits the Internet Age. People who once were primarily consumers of creative works now have the capacity to control them.⁴

But these changes in the legal and technological environment have contributed to a contemporary copyright system that is nearly impossible to navigate. When individual creators claim, retain, and manage their own copyrights, they contribute to a troubling phenomenon I have called “copyright atomism”: the proliferation, distribution, and fragmentation of the exclusive rights bestowed by copyright law. An atomistic copyright system is crowded with protected works and rights, owned by rights-holders who are numerous and far-flung, and whose preferences may be idiosyncratic. This situation raises information and transaction costs for participants in a creative marketplace that is full of works to which copyright presumably applies but for which there are no publicly available indications of who owns the copyright, where the owner is, when copyright will expire, and on what terms the works might be licensed.⁵

The history of atomism in copyright, and the broader history and literature about atomistic property rights more generally, suggest a number of different approaches to the challenges that atomism poses. Some of these approaches aim to limit atomism itself—that is, they structure property rights “holistically” so as to restrict, consolidate, unify, and standardize them. An example from copyright law is the work-for-hire doctrine, which avoids the

3. See generally Van Houweling, *supra* note 2, at 1541 n.26 and accompanying text.

4. See generally Van Houweling, *supra* note 1, at 552–53 (observing that “[i]ndividual creators are . . . increasingly harnessing copyright themselves, insisting on ownership of their rights and controlling the ways in which those rights are licensed to others”).

5. The challenges that atomism poses for the digital copyright environment have also been recognized in other property contexts. For example, Robert Ellickson has explained how transaction costs arise in the tangible property context “from the proliferation of boundaries and ownership entities.” Michael Heller’s work has drawn attention to the resource underuse that can arise when property rights fragment into an “anticommons.” Thomas Merrill and Henry Smith focus on the information cost externalities imposed by the proliferation of customized and idiosyncratic property rights. Clarisa Long, Justin Hughes, Henry Smith, and others have applied this transaction and information cost analysis to intellectual property. In my own work, I have examined how these costs may be imposed in the intangible property context by emerging intellectual property licensing practices. For a fuller discussion and citations to the relevant literature, see Van Houweling, *supra* note 1, at 553–54.

consequences of widely-distributed ownership by individual human authors through the legal fiction of employer authorship. This doctrine thus consolidates ownership of contributions to collective projects undertaken in the employment context (and in some cases of works commissioned from independent contractors).⁶

These anti-atomism approaches are often unpopular with individual authors—increasingly so in an age in which the economics of digital creativity make it possible for individuals to create and publish on their own.⁷ More promising in the digital age, then, may be approaches that aim not to maintain a “holistic” (as opposed to atomistic) copyright structure by consolidating ownership, but rather those that address in a more targeted way the information and transaction costs associated with atomism. Copyright history offers examples of this type of approach as well. For example, early copyright law featured systems of registration and other formalities that produced centralized records of copyright ownership, renewal, and transfer—much like the land records that help to facilitate tracing and transactions in the real property context. Contemporary copyright law has largely abandoned this model, however. Registration requirements and other strict “formalities” are disfavored as prerequisites for copyright protection under national copyright laws and international treaties.⁸

Many observers have lamented the demise of formalities in contemporary copyright law and proposed various ways of “re-formalizing” copyright.⁹ These proposals often do not rely on mandatory compliance or on centralized registries, however.¹⁰ Instead, proposals for (and experiments with) other information-based approaches to the problems posed by atomistic copyright are emerging.

One intriguing possibility is that, in place of (or perhaps in conjunction with) centralized registries, information costs could be reduced with machine-readable tags that attach to and travel with digital works. Copyright law has long recognized this idea of decentralized tagging. One of the previously-mandatory formalities under U.S. law was the provision of copyright notice on published works. Indeed, this requirement was more

6. On the work-for-hire doctrine as an anti-atomism device, *see generally* Van Houweling, *supra* note 1, at 557–58, 594–95, 606.

7. *See generally* Van Houweling, *supra* note 1, at 613–21.

8. On the history and contemporary abandonment of copyright formalities, *see* Christopher Sprigman, *Reform(aliz)ing Copyright Law*, 57 STAN. L. REV. 485 (2004).

9. *See, e.g., id.*; Samuelson et al., *supra* note 2.

10. *See id.*

important than registration.¹¹ Today, provision of notice is not required.¹² But contemporary U.S. copyright law does acknowledge, in a more limited way, the value of decentralized notice provision. It incentivizes notice by making additional remedies available where notice has been provided.¹³ And it also forbids (under some circumstances) the removal of copyright tags¹⁴ (a provision that reflects the WIPO Copyright Treaty's protection of "electronic rights management information").¹⁵

This Article begins to explore the technological and legal environment surrounding decentralized information-based interventions in the atomistic copyright system. As a case study, I will focus on the use of copyright-related, machine-readable metadata embedded in digital photographs.

III. PHOTOGRAPHS AS ATOMISTIC COPYRIGHTED WORKS

Photographs are a key element of many of today's most widely-used platforms for user-generated content and social networking. They are typically the work of individual creators, operating independently and retaining ownership of their copyrights.¹⁶ The proliferation of digital

11. In the contemporary period, registration was required under U.S. law only upon renewal. *See generally* Sprigman, *supra* note 8.

12. *See* 17 U.S.C. § 401 (2010) (allowing but not requiring notice).

13. *See id.* §§ 401(d), 504(c).

14. *See id.* § 1202. *See generally* Susuk Lim, *A Survey of the DMCA's Copyright Management Information Protections: The DCMA's CMI Landscape After All Headline News and McClatchey*, 6 WASH. J.L. TECH. & ARTS 297 (2011); Jane C. Ginsburg, *The Right to Claim Authorship in U.S. Copyright and Trademarks Law*, 41 HOUS. L. REV. 263, 283–84 (2004).

15. World Intellectual Property Organization: Copyright Treaty, Dec. 20, 1996, 36 I.L.M. 65 (1997).

16. The most popular platforms (e.g., Flickr, Facebook, Google) specify in their terms of use that contributors retain copyright in their contributions while granting non-exclusive licenses to the platform operators. For Flickr, operated by Yahoo!, see *Yahoo! Terms of Service*, § 9, YAHOO!, <http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html> (last visited June 27, 2012) ("Yahoo! does not claim ownership of Content you submit or make available for inclusion on the Yahoo! Services. However, with respect to Content you submit or make available for inclusion on publicly accessible areas of the Yahoo! Services, you grant Yahoo! the following worldwide, royalty-free and non-exclusive license(s)"); for Facebook, see *Facebook Statement of Rights and Responsibilities*, § 2, FACEBOOK, <http://www.facebook.com/legal/terms> (last visited June 27, 2012) ("You own all of the content and information you post on Facebook For content that is covered by intellectual property rights . . . you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license"); for Google, including Picasa Web Albums, see *Google Terms of Service*, GOOGLE, <http://www.google.com/intl/en/policies/terms/> (last visited June 27, 2012) ("Some of our Services allow you to submit content. You retain ownership of any intellectual property rights that you hold in that content. . . . When you upload or otherwise submit content to our Services, you give Google (and those we work with) a worldwide license").

photographs with globally scattered individual owners is thus a prime example of atomism in the digital copyright environment. And this atomism has produced concern that information costs could make it difficult to comply with and transact over copyrights in these works.

This anxiety has been most emphatically expressed by professional photographers, who worry that their works will become “orphans” over which they have no effective control.¹⁷ In contrast, other Internet users who share their photographs online do not appear to be motivated by or concerned about copyright and the control it bestows upon them, although copyright nonetheless applies by default to their works. The heterogeneity of both the motivations and the legal and technical sophistication of contributors to this particular cultural environment make it especially difficult to navigate.

Photographs have been recognized as copyrightable subject matter under the U.S. Copyright Act since 1865.¹⁸ Initially there was some question (as there has been in other countries) about whether photographs that aim to capture reality qualify as original works of authorship.¹⁹ But in light of the minimalistic way in which the originality requirement has been interpreted

17. See, e.g., David Riecks, Rivaldo Does & Betsy Reid, *A Metadata Manifesto*, STOCK ARTISTS ALLIANCE (July 2006), <http://www.stockartistsalliance.org/metadata-manifesto-1>.

Lack of information about an image file can delay projects, necessitating additional research to establish licensing rights, obtain clearances, and confirm caption details. This in turn has contributed to the growing problem of misuse of images, whether through error or by intent. Without proper licensing or permissions, users infringe copyright and expose themselves to liability.

Id.

18. See *Goldstein v. California*, 412 U.S. 546, 562 n.17 (1973) (tracing expansion of copyrightable subject matter).

19. See MELVILLE B. NIMMER & DAVID NIMMER, *NIMMER ON COPYRIGHT* § 2.08[E] (2012).

Since its advent in the nineteenth century, photography has posed a basic dilemma for copyright. A rich literature surveys the terrain, discussing whether a photograph simply captures an uncopyrightable fact or qualifies as a work of authorship. The upshot of that debate is that photographs have historically received inferior protection under the copyright laws of other nations Be that foreign treatment as it may, the U.S. experience . . . has embraced photographs as full-fledged members of the copyright genus.

Id. As Nimmer explains, an important early case was *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884), in which the Supreme Court held that a photograph of Oscar Wilde was copyrightable due to the creative choices (pose, lighting, etc.) made by the photographer. But the Court left open the question of whether “the ordinary production of a photograph” qualifies as original.

under U.S. law, the prevailing view is that almost any photograph “may claim the necessary originality to support a copyright merely by virtue of the photographers’ personal choice of subject matter, angle of photograph, lighting, and determination of the precise time when the photograph is to be taken.”²⁰

Despite this statutory and judicial recognition of their copyrightability, many photographs—especially those by individual amateur photographers—were not subject to copyright protection under U.S. law during photography’s first century. Before the 1976 Copyright Act took effect, publication with notice was a prerequisite for protection that many individual photographers did not satisfy. And even those photos that were initially copyrighted were released into the public domain when their owners failed (as most did) to register their renewals.²¹

This situation changed with the 1976 Act, which granted federal copyright protection to original works upon mere fixation, rather than upon publication with notice.²² In addition, the 1976 Act and subsequent legislation eventually did away with the notice requirement even for published works. And the renewal registration requirement became irrelevant because the dual-term system of protection (initial term subject to extension upon renewal registration) was abandoned and replaced with a unitary term

20. Nimmer & Nimmer, *supra* note 19, § 2.08[E]; see also Justin Hughes, *The Photographer’s Copyright—Photograph as Art, Photograph as Database*, 25 HARV. J.L. & TECH. 327, 356 (2012) (“[C]opyright initially struggled with the technology of photography, first came to see the most composed photographs as original art, and then came to reflect artistic theory in accepting that human aesthetic choices permeate many photographs.”); *id.* at 361 (observing the “widespread belief that all photographs are protected by United States copyright law”). *But see id.* at 362 (making the provocative argument that “a large percentage of the world’s photographs are likely not protected by American copyright law because the images lack even a modicum of creativity” and “[i]ndeed, as digitization makes photography more and more ubiquitous, we have probably already crossed a threshold beyond which most of the world’s photographic images are not truly protected by copyright”).

21. See William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 505–06 (2003) (documenting renewal rate of approximately 3% for graphic-arts works, including photographs).

22. The 1976 Act introduced the standard that continues to govern today: “Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.” 17 U.S.C. § 102(a) (2010). “A work is ‘fixed’ in a tangible medium of expression when its embodiment in a copy or phonorecord, by or under the authority of the author, is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration.” *Id.* § 101.

calculated based on the life of the author plus fifty (subsequently extended to seventy) years.²³

As the U.S. Copyright Office noted in its 2006 Report on Orphan Works, these developments made it much more difficult to determine the copyright status and owner of protectable works—especially photographs. As the Report explains in its summary of public comments on the topic:

A search for the owner of copyright in a work almost always begins with information available on the work itself. The comments, however, describe numerous situations involving works that bear no information about the author or the owner of copyright in the work—no name of the author, no copyright notice, no title in short, no indicia of ownership on a particular copy of the work *at all*. . . . The comments show that this obstacle is most pervasive—by far—with photographs. Again and again, the comments point to situations involving inadequate information about the author or owner on individual photographs. Numerous individuals complained about situations where they could not use photographs, or did so with trepidation, because they simply had no way of even knowing who took the picture.²⁴

Examples in the Report illustrate that the problem of orphan photographs is not a new problem of the digital era. It has arisen, for example, when people bring copies of old family photos to professional photo finishers, who refuse to make additional copies without proof of copyright ownership or permission.²⁵ The advent of digital photography and web-based photo sharing introduces new challenges, but new solutions as well—as the next Section describes.

23. 17 U.S.C. § 302(a) (2010) (setting duration for works created on or after January 1, 1978, at “a term consisting of the life of the author and 70 years after the author’s death”). See generally Sprigman, *supra* note 8, at 487–88 (documenting removal of formalities requirements under U.S. law).

24. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 23–24 (2006).

25. *Id.* at 24 (“The most common recurring situation with photographs typically involves the reprinting of old family photographs for preservation, sentimental, or nostalgic purposes. Usually the commenter owns an old, damaged, or deteriorating photograph with no identifying information about the photographer, or outdated information at best. The commenter presents the photographs to a photo-finisher for reproduction, but the finisher refuses to reproduce the work.”).

IV. STRATEGIES FOR MITIGATING THE COSTS OF ATOMISTIC OWNERSHIP

A. REINVIGORATING REGISTRATION

One proposed solution to the orphan works dilemma and other problems posed by atomistic ownership of photographs and other copyrighted works is to reinvigorate the registration system—perhaps with reliance on private registries instead of or in addition to the records maintained by the Copyright Office. For example, the Copyright Principles Project has recommended that “[c]opyright law should encourage copyright owners to register their works so that better information will be available as to who claims copyright ownership in which works,”²⁶ and also that the Copyright Office “should transition away from being the sole registry for copyrighted works and toward certifying the operation of registries operated by third parties, both public and private.”²⁷

The registry solution—whether publicly or privately maintained²⁸—seems less promising for photographs and other images than for some other types of works, however. A text-based registry may not be very useful for someone trying to establish the copyright status of an apparently orphaned photograph, for example. As the Report on Orphan Works explains:

Very often any description of a typical orphan work photograph would be useless: for example, “farmer with three horses, mountains in background.” A user might never match the photograph to the description. Conversely, requiring an owner to monitor a user database would be pointless for the same reason. In addition, adding thumbnail images of the photograph might not help either (even if it were legal and logistically feasible) because the database of images might quickly swell to millions of images: owners and users could still miss each other in the mix.²⁹

Technological advances may make image registries more useful in the future. There have been relevant innovations even in the few years since the Report on Orphan Works was issued. For example, TinEye—a service that launched in May 2008—describes itself as “a reverse image search engine. You can submit an image to TinEye to find out where it came from, how it is being used, if modified versions of the image exist, or to find higher

26. Samuelson et al., *supra* note 2, at 1198 (“Recommendation #1”).

27. *Id.* at 1203 (“Recommendation #2”).

28. See generally Peter S. Menell, *Knowledge Accessibility and Preservation Policy for the Digital Age*, 44 HOUS. L. REV. 1013 (2007) (discussing the importance of public oversight of accessibility and preservation).

29. REPORT ON ORPHAN WORKS, *supra* note 24, at 75 (citations omitted).

resolution versions.”³⁰ TinEye claims to be “the first image search engine on the web to use image identification technology rather than keywords, metadata or watermarks.”³¹ In 2011 Google launched its own reverse image search functionality.³² These services do not purport to provide any information about copyright, but similar technology could be used to make a copyright registry searchable by someone in possession of an image of unknown provenance.

B. DIGITIZING DECENTRALIZED NOTICE

Other technological developments may facilitate a different approach to the information challenges of atomistic copyright in photographs: decentralized notice provision through tagging of the image files themselves (including tagging that refers to a text-based registry or other source of additional information). The Report on Orphan Works notes the importance of the decentralized notice approach:

For authors and copyright owners, marking copies of their works with identifying information is likely the most significant step they can take to avoid the work falling into the orphan works category. This is particularly true for works of visual art, like photographs and illustrations, that otherwise do not contain text or other information that a user can rely on to help determine the identity of the copyright owner.³³

As noted above, this type of marking has long been a part of the U.S. Copyright system, which used to require copyright notice (e.g. “Copyright © 2011 University of California Press”) as a prerequisite for protection of published works.³⁴ Although notice is now voluntary, technological developments may make notice easier for copyright holders to provide, easier for users to find and understand, and therefore more ubiquitous and useful. Relevant technological innovations include techniques and standards for

30. *About*, TINEYE, <http://www.tineye.com/about> (last visited July 31, 2012).

31. *Frequently Asked Questions*, TINEYE, <http://www.tineye.com/faq#what> (last visited June 27, 2012).

32. *Google Search by Image*, GOOGLE, <http://www.google.com/insidesearch/features/images/searchbyimage.html> (last visited June 27, 2012); see also Alex Wild, *Google's Reverse Image Search*, SCIENTIFIC AMERICAN BLOG NETWORK (Aug. 16, 2011), <http://blogs.scientificamerican.com/compound-eye/2011/08/16/googles-reverse-image-search/>.

33. REPORT ON ORPHAN WORKS, *supra* note 24, at 99.

34. See Sprigman, *supra* note 8, at 493–94 (describing evolution of notice requirement under the U.S. Copyright Act).

embedding copyright-related “metadata” in digital files.³⁵ A consortium of digital media companies called the Metadata Working Group³⁶ (“MWG”) is among the advocates for effective use of metadata stored in digital files. The group’s Guidelines for Handling Image Metadata offer this helpful explanation:

Metadata, often referred to as “data about data,” provides interesting information that supplements the primary content of digital documents. Metadata has become a powerful tool to organize and search through the growing libraries of image, audio and video content that users are producing and consuming. This is especially important in the area of digital photography where, despite the increased quality and quantity of sensor elements, it is not currently practical to organize and query images based only on the millions of image pixels. Instead, it is best to use metadata properties that describe what a photo represents and where, when, and how the image was taken.³⁷

The MWG Guidelines identify several key metadata fields, the most familiar of which include:

- keywords (referred to as “tags” in many photo applications)³⁸
- description (often referred to as “caption”)³⁹
- date/time (including when the photo was taken, when it was digitized, and when it was modified)⁴⁰
- location (both location created and location depicted)⁴¹
- creator (or “author”)⁴²
- copyright (specifically, copyright notice and a URL to more information about the copyright).⁴³

35. See generally NAT’L INFO. STANDARDS ORG. UNDERSTANDING METADATA (2004), available at <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>.

36. See METADATA WORKING GROUP, <http://www.metadataworkinggroup.org/> (last visited June 27, 2012) (describing its goals as “[p]reservation and seamless interoperability of digital image metadata” and “[i]nteroperability and availability to all applications, devices, and services”); *Member Companies*, METADATA WORKING GROUP, <http://www.metadataworkinggroup.org/members> (last visited June 27, 2012) (identifying member companies Adobe, Apple, Canon, Microsoft, Nokia, and Sony Corporation).

37. METADATA WORKING GROUP, GUIDELINES FOR HANDLING IMAGE METADATA 8 (Version 2.0, Nov. 2010) [hereinafter MWG GUIDELINES], available at http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf.

38. *Id.* at 35.

39. *Id.* at 36.

40. *Id.* at 37.

41. *Id.* at 45.

42. *Id.* at 43.

Some of this metadata can be automatically embedded in digital files when they are initially created.⁴⁴ Many modern digital cameras mark photographs with metadata including date, time, and location, for example. Other types of metadata—including copyright information as well as descriptions of the subject matter and other contextual information—can be added manually by the photographer, either photograph-by-photograph or in bulk. Some capture devices (many smartphones, for example) have interfaces that allow manual input of this type of information pre- and/or post-capture. Metadata can also be added using photo-editing software after an image has been loaded onto a computer. And some web-based photo hosting platforms allow users to add additional metadata to hosted photos (during the uploading process and/or afterwards).⁴⁵

The metadata vision that is expressed in the MWG Guidelines is one of interoperability and persistence across the various devices and applications that handle digital images. The ideal is for metadata associated with an image to be durably embedded in the image file in a standards-compliant way that will make it recognizable and usable by subsequent machines, software programs, and ultimately by humans for whom this information is valuable.⁴⁶

In addition to the MWG, organizations representing independent photographers have also taken a leading role in advocating what they see as the proper handling of metadata. In 2006, the Stock Artists Alliance (“SAA”)⁴⁷ released “A Metadata Manifesto,” which insists that “[m]etadata is

43. *Id.* at 42.

44. See Peter Krough, *Metadata Overview*, DPBESTFLOW.ORG, <http://www.dpbestflow.org/metadata/metadata-overview> (last visited June 27, 2012) (including an explanation of “[h]ow does metadata get there?”).

45. See, e.g., *Add Geotags in Picasa*, GOOGLE, <http://support.google.com/picasa/bin/answer.py?hl=en&answer=161869&topic=1689810&ctx=topic> (last visited June 27, 2012).

46. The Guidelines observe—and attempt to rectify—obstacles to this vision. “Different applications and devices have chosen to handle ambiguously or weakly-defined metadata specifications in different ways.” MWG GUIDELINES, *supra* note 37, at 10. “Different applications and devices have chosen to follow different policies in cases where metadata can be stored in more than one standard location.” *Id.* “An application or device often stores proprietary metadata, such as maker notes, within a metadata container. This practice is fragile because such private data can easily be lost when a different application modifies a file.” *Id.* “Some applications and devices usurp general purpose metadata properties to address specific needs. This can cause compatibility problems for applications that correctly use those properties in accordance with the generally accepted specification.” *Id.* “Some applications avoid the complexities of storing metadata within image files altogether and opt, instead, to store it in a separate file or database. This practice can easily result in the loss of metadata when a file is used across several applications.” *Id.*

47. “SAA is an international association of professional photographers and others who create visual works for stock licensing. The mission of SAA is to support and protect the business interests of professional stock photographers worldwide.” *The SAA Mission*

essential to identify and track digital images.”⁴⁸ The document goes on to explain:

Everyone involved with digital images needs to recognize that embedded metadata is an essential part of every digital image. We feel it is the most efficient means of storing information about an image, and the best way to ensure that users can easily find the image and identify its source after it has been distributed or exchanged.⁴⁹

The SAA Manifesto emphasizes, in particular, the role that embedded metadata could play in solving the challenges that atomistic copyright poses both for creators and users—especially in the digital age:

The volume of digital files challenges publishers who need to manage and access them. Busy designers and art directors download preview images to their desktops, only to find weeks or months later they cannot identify the source. Librarians and curators—charged with making more cultural resources available to the public—are already overburdened managing their legacy analog material. Now, they must cope with rapidly expanding digital assets as well. Lack of information about an image file can delay projects, necessitating additional research to establish licensing rights, obtain clearances, and confirm caption details. This in turn has contributed to the growing problem of misuse of images, whether through error or by intent. Without proper licensing or permissions, users infringe copyright and expose themselves to liability. The pressures on image creators—as copyright holders—to protect their intellectual property has intensified since the digitalization and online distribution of their images. If their images cannot be properly identified, they suffer from lost revenues due to missed licensing opportunities.⁵⁰

Among the principles the SAA Manifesto proposes to address these challenges is that “[o]wnership metadata must never be removed.”⁵¹

Similarly, the International Press Telecommunications Council (“IPTC”), a membership organization that develops standards for exchange of content

Statement, STOCK ARTISTS ALLIANCE, <http://www.stockartistsalliance.org/info/about/mission> (last visited June 27, 2012).

48. Riecks et al., *supra* note 17, at 3.

49. *Id.* at 4.

50. *Id.* at 2.

51. *Id.* at 3. The other principles are: “Metadata is essential to identify and track digital images.” *Id.* “Metadata must be written in formats that are understood by all.” *Id.*

(including photographs) within the news industry,⁵² has issued its own “Embedded Metadata Manifesto,” which stresses the importance of persistent and interoperable embedded metadata. The IPTC Manifesto devotes particular attention to copyright-related metadata.⁵³

Like the SAA, the IPTC is concerned about the fate of orphaned photographs in the digital age. Its Manifesto observes that “[p]hotographers, film makers, videographers, illustrators, publishers, advertisers, designers, art directors, picture editors, librarians and curators all share the same problem: struggling to track rapidly expanding collections of digital media assets such as photos and video/film clips.”⁵⁴ To address the copyright aspects of this problem, the IPTC Manifesto insists that “[c]opyright management information metadata must never be removed from the files” because “[o]wnership metadata is the only way to save digital content from being considered orphaned work.”⁵⁵

These manifestos suggest that embedded metadata could address some of the challenges that I associate with atomistic copyright, by reinvigorating the copyright notice function that was long a part of U.S. copyright law. A photographer could embed all of her photos with her name, copyright notice, and a link to license terms and contact information, all in a format that reliably traveled with her digital images across multiple digital platforms—from her camera to her editing software to a web-based platform.⁵⁶ Unlike

52. *IPTC Membership Questions and Answers*, INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL (2009), available at http://www.iptc.org/download/public/IPTC-Membership-QandA_2009.1.pdf (“The International Press Telecommunications Council is a consortium of news agencies, newspapers and news system vendors that develops and publishes technical specifications to promote the easy, accurate and inexpensive sharing of news.”).

53. *Embedded Metadata Manifesto (2011)*, INTERNATIONAL PRESS TELECOMMUNICATIONS COUNCIL, [http://www.iptc.org/site/Photo_Metadata/Embedded_Metadata_Manifesto_\(2011\)](http://www.iptc.org/site/Photo_Metadata/Embedded_Metadata_Manifesto_(2011)) (last visited June 27, 2012).

54. *Id.*

55. *Id.*; see also *FAQ: ASMP's Position on Orphan Works*, AMERICAN SOCIETY FOR MEDIA PHOTOGRAPHERS, <http://asmp.org/articles/frequently-asked-questions-about-asmps-position-orphan-works.html> (“[Y]ou should be embedding metadata and communicating with clients to keep that data from being stripped. We are working with software and hardware makers to secure metadata as best we can.”).

56. See generally Abigail L. Dansiger, *Embedded Metadata: Friend or Foe to Our Digital Collections?*, *Library Student Journal* (Jan. 2011), <http://www.librarystudentjournal.org/index.php/ljsj/article/viewArticle/210/290>.

[E]mbedding metadata in images is receiving attention from various parties, such as digital asset managers, professional photographers, federal agencies, and lawyers. These groups have identified embedding metadata as the best way to ensure copyright protection and avoid future litigation

mere textual notices, machine-readable notices could trigger features (pop-up alerts about licensing terms, for example) that would help to ensure that users in fact paid attention and complied with copyright holders' preferences where appropriate. Embedded metadata could also include links to registries that would provide additional information, thus combining decentralized tags with centralized resources like that proposed by the Picture Licensing Universal System ("PLUS"),⁵⁷ which is now testing a registry that it promises will ultimately be used "to find rights and descriptive information . . . for any image, and to find current contact information for related creators, rights holders and institutions."⁵⁸

This vision of a seamless and ubiquitous system of copyright-related (and other) metadata connected to digital image files has not yet been fully realized, however. Indeed, the MWG Guidelines and these metadata manifestos represent attempts to improve metadata-handling practices that are seen as inadequate. In particular, metadata enthusiasts complain that some devices and applications do not embed metadata in a durable and standards-compliant way;⁵⁹ others strip metadata from incoming files;⁶⁰ and many handle metadata in confusing and inconsistent ways that make it difficult for photographers and photo users to understand.⁶¹ Meanwhile, metadata skeptics point out that the best practices for handling embedded

issues or orphan works, which result when copyright holders cannot be identified.

Id.

57. See Riecks et al., *supra* note 17, at 3.

Imagine a world where metadata is ubiquitous. It's a world where images can be easily located and identified by anyone, anywhere. Creators can transmit their images to distributors and users, who instantly integrate these into their systems. Image users can track their digital assets using fully automated systems. A registry—now in development by the Picture Licensing Universal System (PLUS)—will link every image to current information about its source and owner.

Id.

58. *About*, PLUS REGISTRY BETA 1.0, <https://www.plusregistry.org/cgi-bin/WebObjects/PlusDB.woa/1/wo/w8sKt5k98zlrjyCVcb39sg/0.99.27> (last visited July 20, 2012).

59. See *supra* note 46 (noting problems identified by the MWG); see also *Embedded Metadata Manifesto*, *supra* note 53 ("To support the exchange of metadata with content it is a business requirement that file formats embed metadata with the digital file. Other methods like sidecar files are potentially exposed to metadata loss.").

60. See, e.g., Riecks et al., *supra* note 17, at 3 ("Stock photographers, for example, provide digital files to single or multiple distributors who commonly omit metadata from the image files they post online.").

61. MWG GUIDELINES, *supra* note 37, at 10 (citing "significant frustration for users who want consistent metadata interoperability across digital imaging products and services").

metadata are no guarantee that the metadata will be substantively accurate.⁶² In the next Part, I survey and assess these challenges.

V. ANXIETY ABOUT AUTOMATION

As a solution to the problems posed by atomistic copyright in digital photographs, providing distributed digital notice via embedded metadata faces several challenges. These can be roughly categorized as relating to (1) interoperability, (2) persistence, (3) user comprehension, and (4) accuracy. There have recently been important developments on all of these fronts. Progress toward interoperability includes the development and adoption of metadata standards and guidelines for inter-standard compatibility, and the emergence of standardized copyright licenses that increase the potential for legal as well as technical interoperability. Persistence and interoperability are related, as metadata loss can result from incompatibilities. But persistence also depends on policy choices, which I discuss below. As for user comprehension, this too is linked to both technical and legal interoperability, because consumers will find it difficult to understand inconsistent and incompatible schemes. The adoption of standardized licensing tools by major photo hosting services promises to help on this front, while also illustrating the as-yet unrealized potential of license-related metadata. Ensuring the provenance and accuracy of copyright and license-related metadata might be the most intractable challenge—one that technology did not create and may be insufficient to solve completely.

A. INTEROPERABILITY

1. *Technical Interoperability*

Many promising uses of metadata require that it be recognized across multiple devices and applications—for example, from the camera in which an image is captured, to the software with which it is edited, to the web platform on which it is hosted, etc. This type of interoperability relies upon standards for how metadata will be written and processed. As of 2006, an Adobe technical paper lamented that “[b]ecause of the lack of coordination between manufacturers and organizing bodies, much of the potential of metadata remains unrecognized.”⁶³ Adobe has put forward its own proposed solution to this problem in the form of the open-source Extensible Metadata

62. See, e.g., Cory Doctorow, *Metacrap: Putting the Torch to Seven Straw-men of the Metatopia*, THE WELL, <http://www.well.com/~doctorow/metacrap.htm>.

63. *Technical Paper: About Metadata*, ADOBE, http://www.adobe.com/digitalimag/pdfs/phscs2ip_metadata.pdf.

Platform (“XMP”), which the company describes as “a standard for the definition, creation, and processing of metadata.”⁶⁴

Adobe has also partnered with Apple, Canon, Microsoft, Nokia, and Sony to participate in the Metadata Working Group (“MWG”), mentioned above, with the goals of “[p]reservation and seamless interoperability of digital image metadata”; “[i]nteroperability and availability of metadata to *all* applications, devices, and services”; and “[e]nabling emerging metadata-based workflows.”⁶⁵

It remains to be seen whether these efforts will succeed. The MWG membership includes producers of digital photography capture devices and of photo-editing software. Some of the members also maintain web-based services that can be used to share photos.⁶⁶ But it is nonetheless a relatively narrow industry group, not a broad-based standards-setting body. Notably, its membership does not currently include the leading hosts of digital photographs, such as Google, Yahoo! (owner of the online photo-management site Flickr), and Facebook. Google’s Picasa now preserves XMP compatible caption metadata,⁶⁷ and Flickr also preserves at least some metadata on photos uploaded to and downloaded from its service. But neither platform appears to have fully embraced the practices promoted by Adobe and its MWG partners. For its part, Facebook appears to remove almost all metadata from photos uploaded there.⁶⁸

64. *XMP Specification Part I*, at 7 (2010), ADOBE, available at <http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart1.pdf>; see also Fed. Digitization Initiative Still Image Working Grp., *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* (Aug. 2010), DIGITIZATION GUIDELINES, http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf (“Adobe’s Extensible Metadata Platform (XMP) is another emerging, standardized format for describing where metadata can be stored and how it can be encoded, thus facilitating exchange of metadata across applications.”).

65. MWG GUIDELINES, *supra* note 37, at 11.

66. See, e.g., *PlayMemories Online*, SONY, <http://playmemoriesonline.com/> (last visited Nov. 30, 2012).

67. As Picasa’s user documentation explains: “When you create captions for your photos, they’re saved to your photos and stay with them, even when you upload them, export them, or send them by email. Your captions will be displayed in Picasa, Picasa Web Albums and other EXIF, IPTC and XMP compatible programs.” *Creating and Editing Captions*, GOOGLE (Sept. 27, 2011), <http://picasa.google.com/support/bin/answer.py?hl=en&answer=15055>; see also *Data Stored in Photos (EXIF)*, GOOGLE (Sept. 27, 2011), <http://picasa.google.com/support/bin/answer.py?hl=en&answer=1208119>.

68. See *The Controlled Vocabulary Survey Regarding the Preservation of Photo Metadata by Social Media Websites*, CONTROLLED VOCABULARY, <http://www.controlledvocabulary.com/social-media/> (last visited June 29, 2012).

Where interoperability efforts succeed, users are able to embed metadata with their capture devices and to upload those photos directly to web hosting services where the metadata is preserved and displayed according to the users' preferences. For example, users of image capture devices are increasingly able (1) to elect (now typically via an opt-in mechanism) whether their image will be tagged with geographic information; and (2) to add captions and other descriptive metadata to their images.⁶⁹ Although far from foolproof or ubiquitous, these metadata at least sometimes stay embedded in the images as they are uploaded onto web platforms.⁷⁰

There is no technical obstacle to using the same functionality to embed and preserve metadata related to the copyright and licensing status of photographs.⁷¹ The MWG Guidelines include recommendations for using XMP to embed both a standard copyright notice and a URL to more information about the copyright.⁷² These recommendations promise to promote the inclusion of copyright-related information in standardized locations within image metadata. But they do not do much to promote the standardization of the content of that metadata beyond referring to the basic elements of copyright notice (copyright, date, and owner). For example, if a user wants to provide a URL to more information, what might that information be? And will it be standardized such that subsequent users will understand the terms under which the photograph may be used? Could all of this be simple enough that it could be implemented via the same types of on-screen interfaces that allow smartphone users to specify their geo-tagging preferences for uploaded photos? While software companies and hardware manufacturers focus on technical interoperability, these questions also raise issues of *legal interoperability*.

69. Nokia camera phones feature tools for tagging photos with descriptions and geotags and then for uploading them directly onto either Flickr or Nokia's Ovi platform. *See, e.g., Nokia N900 User Guide, Share Your Images and Videos*, NOKIA, at 76, 82, available at http://www.nokia.com/us-en/support/product/nokia-n900/userguide/?action=onlineuserguidepagechange&pFile=GUID-03220374-A1AE4393-8F2F-D79A29D18A79_FILE001.html (2010). The iPhone Flickr app invites users to add a title, description, tags, and sets. It also asks whether the user wants to share the location as identified by a GPS-enabled phone. The iPhone interface also asks whether the photo should be public or private. *See Flickr for iPhone*, YAHOO!, <http://mobile.yahoo.com/flickr/iphone> (last visited June 29, 2012). Users of Android phones can upload photos directly to Picasa Web Albums and add a caption. Other phones can access the mobile Picasa site via a web browser and upload to it. *See Uploading on Android*, GOOGLE, <http://support.google.com/mobile/bin/answer.py?hl=en&answer=132550>.

70. *See The Controlled Vocabulary Survey*, *supra* note 68.

71. *See* MWG GUIDELINES, *supra* note 37, at 42 (addressing copyright metadata).

72. *Id.*

2. *Legal Interoperability*

One organization working on issues of legal interoperability is Creative Commons.⁷³ This non-profit “that enables the sharing and use of creativity and knowledge through free legal tools”⁷⁴ has promulgated a menu of public licenses that can be associated with copyrighted works (“CC licenses”), permitting reuse of those works under specified circumstances. CC licenses are “written to conform to international treaties governing copyright.” They are “all intended to be effective anywhere in the world, with the same legal effect.”⁷⁵ There are six basic CC licenses applicable to copyrighted photographs. Licensing information that corresponds to these six standardized choices is increasingly becoming a familiar component of web-based photo hosting platforms.

The photo-sharing platform Flickr has offered its users the option of applying CC licenses to their photos since at least 2005. When a Flickr user uploads a photo from her computer, she is invited to add a title, description, and tags. She can create sets and collections in which to organize this and other photos. She can also edit the “owner settings,” including copyright information, specifying either “all rights reserved,” or “some rights reserved.” The “some rights reserved” accompanies images for which the uploading user has elected to apply a CC license to the image, thereby inviting copying and other reuses under specified circumstances. Instead of making this selection for each photo, Flickr users may also select a copyright option as the default for all of their photos. To date, close to two-hundred million Flickr photos have been licensed on CC terms.⁷⁶

More recently, Google’s Picasa Web Albums began to offer the CC licensing option as well. By default, photos uploaded to Picasa Web Albums are displayed with a copyright symbol and the words “all rights reserved.” But this copyright notice can be modified at the option of the user. When viewing her own album using the web interface, a user sees an “edit” link beside the words “All rights reserved.” Clicking “edit” presents the user with the options “Do not allow reuse” or “Allow reuse.” If “Allow reuse” is selected, then the user has the further option of allowing remixing, allowing

73. As acknowledged above, I serve on the Board of Directors of Creative Commons. I also served on its staff from 2001 to 2002.

74. *About*, CREATIVE COMMONS (Aug. 3, 2012, 3:45 PM), <http://creativecommons.org/about>.

75. *CC Affiliate Network: The Licensing Suite*, CREATIVE COMMONS (Mar. 15, 2012, 1:15 PM), http://wiki.creativecommons.org/CC_Affiliate_Network.

76. *Explore/Creative Commons*, FLICKR, <http://www.flickr.com/creativecommons/> (last visited June 29, 2012).

commercial use, and requiring “share-alike” (various combinations of which yield the six basic Creative Commons licenses). Depending on the selection, the photo is subsequently displayed with the words “all rights reserved” or “some rights reserved” (and corresponding icons that represent the CC license choice). A Picasa Web Album user can also apply a Creative Commons license by default to all of her photos by making a selection within the “privacy and permissions” settings of her album.⁷⁷

There remains incompatibility between some of the various CC license types,⁷⁸ and between CC licensing and other terms under which copyright holders might choose to make their photographs available.⁷⁹ But the limited CC menu and its increasing prominence as an option for users of web-based photo platforms promise to deliver to this environment a new degree of legal interoperability—by which I mean standardization of license terms that facilitates comprehension by licensors and licensees, and that facilitates reuse by increasing the probability that multiple resources will be available under the same terms and thus amenable to combined reuse.

Creative Commons aims to link this legal interoperability with metadata interoperability. The licenses are designed to correspond to metadata that can be embedded in digital files, and Creative Commons has developed a metadata schema that is compatible with XMP.⁸⁰ The Picture Licensing Universal System (“PLUS”) has also developed an XMP-compatible “License Data Format,” described as “a metadata ‘schema’ for an embedded image license.”⁸¹

Combining these existing elements would allow photographers to use a simple interface on their capture devices to select copyright licensing options,

77. *See Changing Licensing Options*, GOOGLE (Dec. 11, 2011), <http://support.google.com/picasa/bin/answer.py?hl=en&answer=94293&topic=1689812&ctx=topic>.

78. *Frequently Asked Questions*, CREATIVE COMMONS (Apr. 23, 2012, 6:54 PM), http://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_combine_two_different_Creative_Commons_licensed_works.3F.

79. *See generally* Van Houweling, *supra* note 1, at 634–35.

80. *XMP*, CREATIVE COMMONS (June 1, 2011, 9:32 PM), <http://wiki.creativecommons.org/XMP>. Based on informal experiments using my own CC-licensed photographs, Flickr and Picasa do not yet appear to embed metadata corresponding to CC license terms—or at least not in ways that are as useful as one would hope. For example, when I downloaded from Flickr a photo that I had posted there with a CC “Attribution” license and then uploaded it to Picasa, its apparent status had reverted to “All Rights Reserved.” Either the license-related metadata was not embedded in the file when I made the selection on Flickr, it was not retained when I uploaded it to Picasa Web Albums, or it was retained but not interpreted and implemented in the way I would expect.

81. *PLUS License Data Format*, PLUS COALITION, INC. (2011), <http://www.useplus.com/useplus/license.asp>.

which could then be embedded into the digital files (along with captions, geographic location, etc.) and retained as those files were loaded onto web-based platforms. Although neither the technical nor legal standards have been universally adopted, and although even adopters have not combined every element of the legal and technical infrastructure I have just described, the progress that has been made illustrates the potential for uniting legally interoperable licensing terms with technically interoperable metadata standards and practices to facilitate the kind of notice provision to which copyright reformers aspire.

B. PERSISTENCE

For metadata to provide useful information about the copyright status, ownership, and license terms of digital images (and other works), it has to stay attached to those images. Photographers who are enthusiastic about the promise of metadata stress this point. For example, the Stock Artists Alliance's 2006 Metadata Manifesto highlights this issue, and its impact on both photographers and image users:

[I]mage creators have no control over what happens to that metadata once their files are circulated. It can be easily removed, as we lack the technological means to create permanent or protected metadata.

Stock photographers, for example, provide digital files to single or multiple distributors who commonly omit metadata from the image files they post online. The problem worsens as files are then forwarded to sub-distributors. Along the way, file names get changed and metadata may be altered or stripped. . . . Typically, there is no metadata in the images, and the stock distributor file names are not helpful for identification. . . .

For image users who are downloading images, archiving, and later repurposing these digital files, the loss of critical information along the way can be detrimental. Without licensing metadata, they cannot determine their rights to use the image. Without contact metadata, they cannot easily inquire about these rights. And without caption metadata, they may not be able to identify who or what is in the image.⁸²

Regarding ownership metadata in particular, the SAA Metadata Manifesto insists that it never be removed without copyright owner authorization:

82. Riecks et al., *supra* note 17, at 3 (emphasis in original).

Ownership metadata must never be removed.

We need to institute standards and best practices in order to protect and preserve critical metadata. Information that identifies the copyright holder must be treated as “read only” or “write once” data, and must never be removed by image distributors and users. The only exception would be changes done with the explicit consent of the copyright owner.

Automated systems for creating and managing digital files need to honor and assist implementation of this principle. Most critically, these systems need to preserve ownership metadata *by default* and discourage removal of other metadata by warning users about the legal implications of removal.⁸³

The Manifesto explains how device interfaces can serve this goal, urging technology providers to “[m]ake preserving metadata the default option when copying, saving or exporting any image” and to “[i]nsure that user interfaces—for operations that remove metadata—are designed to allow users to clearly understand what they are about to do.”⁸⁴

The Metadata Working Group’s Guidelines seem to recognize without fully tackling difficult questions regarding the removal and/or alternation of metadata. The Guidelines articulate rules for “changer applications,” that “read metadata from an image file and then write[] new or modified metadata back to a file.”⁸⁵ Photo sharing platforms like Facebook and Flickr seem to fall within this definition to the extent they modify metadata embedded in uploaded user photos. Section 3.1.2 of the Guidelines addresses removal of metadata by changer applications: “Deletion of metadata MUST only be done with specific intent.”⁸⁶ More specifically: “Descriptive metadata, information added by a user, must only be deleted by explicit user intent. Non-descriptive metadata may however be deleted with explicit intent. It should only be modified or deleted if it is known to be inaccurate or problematic.”⁸⁷

This language is unclear on several points. What counts as “explicit user intent” required to delete descriptive metadata? And as for non-descriptive metadata, *whose* explicit intent is required? And *who* should have knowledge that the data is “inaccurate or problematic”? The Guidelines’ Implementation Notes acknowledge this lack of precision:

83. *Id.* at 4 (emphasis in original).

84. *Id.* at 5.

85. MWG GUIDELINES, *supra* note 37, at 16.

86. *Id.*

87. *Id.*

The phrases “done with specific intent” or “by explicit user intent” are intentionally vague. There can be many kinds of intent. For example, a user might ask that a saved file be as small as possible, or be redacted for publication. It is a legitimate design choice for this to involve deletion of metadata. Ideally additional safeguards should be placed around sensitive metadata such as a copyright, but that is the purview of the application, not the MWG.⁸⁸

Although outside the purview of the MWG, this question of when, if ever, copyright information and other metadata should be removed as an image file moves from one platform to another is clearly an important one to some users—and also to a copyright system faced with orphan works and other problems associated with atomistic copyright.

As a legal matter, this type of metadata is protected against removal and alteration under limited circumstances. Provisions of U.S. copyright law added by the Digital Millennium Copyright Act (“DMCA”) address “copyright management information,” which is defined to include specified categories of “information conveyed in connection with copies or phonorecords of a work or performances or displays of a work, including in digital form.”⁸⁹ The covered types of information include, *inter alia*, the author, copyright owner, and terms and conditions for use of the work.⁹⁰

The statute prohibits the provision of false copyright management information and the unauthorized removal or alteration of copyright management information.⁹¹ But both prohibitions also include an intent element. Provision of false information is prohibited when knowing and “with the intent to induce, enable, facilitate, or conceal infringement.”⁹² Removal or alteration is prohibited by someone “knowing, or . . . having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement.”⁹³ This provision seems to recognize the importance of copyright-related metadata and to preserve it against egregious abuses, but it is unlikely that the removal of copyright-related metadata that often occurs when digital images are transferred from one platform to another will violate the prohibition on removal of copyright management information due to the lack of culpable knowledge or intent.

There is, however, another provision of the DMCA that may eventually be more useful for protecting metadata. Section 512 limits the liability of

88. *Id.* at 71.

89. 17 U.S.C. § 1202(c) (2010).

90. *Id.*

91. *Id.* § 1201(a)–(b).

92. *Id.* § 1202.

93. *Id.*

online service providers (“OSPs”) that store material on behalf of their users (and perform other network functions) under specified circumstances.⁹⁴ Widely referred to as the “OSP safe harbor,” this provision imposes various eligibility requirements, including that a covered OSP “accommodates and does not interfere with standard technical measures.”⁹⁵ “Standard technical measures” are defined as:

technical measures that are used by copyright owners to identify or protect copyrighted works and—

(A) have been developed pursuant to a broad consensus of copyright owners and service providers in an open, fair, voluntary, multi-industry standards process;

(B) are available to any person on reasonable and nondiscriminatory terms; and

(C) do not impose substantial costs on service providers or substantial burdens on their systems or networks.⁹⁶

To date, it appears that device manufacturers and software companies, along with some specialized communities of copyright holders and public licensing advocates like Creative Commons, have been most actively involved in developing standards for embedding metadata in digital images and other types of files. This probably does not yet represent the “broad consensus of copyright owners and services providers” to which the statute refers. That standard may be satisfied as these efforts move forward, however.⁹⁷ If it is, social networking services and other platforms used to distribute photographs will have a strong incentive to “accommodate” and “not interfere” with embedded metadata, which for some of them would

94. *Id.* § 512.

95. *Id.* § 512(i)(1)(B).

96. *Id.* § 512(i)(2).

97. See generally Jonathan J. Darrow & Gerald R. Ferrera, *Social Networking Web Sites and the DMCA: A Safe-Harbor From Copyright Infringement Liability or the Perfect Storm?*, 6 NW. J. TECH. & INTELL. PROP. 1, 17 (2007) (“[A]lthough it is likely that watermarks and copyright management systems will *one day* fall into the category of standard technical measures, they do not do so today.”); Lauren G. Gallo, *The (Im)possibility of “Standard Technical Measures” for UGC Websites*, 34 COLUM. J.L. & ARTS 283, 300 (2011) (reviewing the relevant case law and legislative history, noting that “the phrase ‘standard technical measures’ has yet to attach to any existing technology”). Cf. Jane C. Ginsburg, *Separating the Sony Sheep from the Grokster Goats: Reckoning the Future Business Plans of Copyright-Dependent Technology Entrepreneurs*, 50 ARIZ. L. REV. 577, 590–91 & n.57 (2008) (concluding that “[a]rguably, filtering technology might be such a measure,” but that “the definition . . . suggests that the present state of filtering technologies may not suffice, principally because there is not yet an inter-industry consensus regarding the design and implementation of filtering measures”).

mean changing their current practices of removing metadata when uploading user files (unless they could establish that doing so would “impose substantial costs . . . or substantial burdens on their systems or networks”).⁹⁸

C. USER COMPREHENSION

Increased adoption of standards for both the handling of embedded metadata and its content and interpretation will also help to address another challenge: ensuring that users understand how their photos are being marked and how that information can be communicated to others. Some metadata practices have produced user confusion and surprise about both the unexpected retention of metadata and its unexpected removal.

The metadata manifestos quoted in Section V.B, *supra*, reflect the frustration of relatively sophisticated, metadata-savvy photographers who are disappointed when photo metadata is removed without their permission. The opposite type of surprise and confusion has been most notable with regard to privacy-oriented objections to “geo-tagging” of photos with location information that can be automatically collected and embedded into photos by GPS-enabled devices. Depending on how it is subsequently handled, this location information can then be displayed publicly along with the photographs. Photographers who do not understand the capacity of their equipment and the photo-sharing applications they use have been surprised by these revelations.⁹⁹ Increasingly—and likely in response to consumer surprise and dissatisfaction—this functionality is disabled by default and users have to affirmatively choose to turn on geo-tagging functionality.¹⁰⁰

These developments demonstrate the interrelatedness of metadata standardization, user comprehension, and privacy. As metadata practices become more uniform and widespread, users are likely to understand them better (or, in some cases, to object to and prompt changes in emerging practices).

98. 17 U.S.C. § 512(i)(2)(C).

99. See, e.g., Kate Murphy, *Web Photos That Reveal Secrets, Like Where You Live*, N.Y. TIMES, Aug. 11, 2010, at B6, available at <http://www.nytimes.com/2010/08/12/technology/personaltech/12basics.html>.

100. See, e.g., *iOS 4: Understanding Location Services*, APPLE (June 11, 2012), <http://support.apple.com/kb/HT1975>.

Location warnings are the requests made by applications (such as Camera, Compass, and Maps as well as location-dependent third-party apps) to use Location Services with those applications. An application will present a location warning the first time it needs to access Location Services data. Tapping OK will give that app permission to use Location Services as needed.

Id.

D. ACCURACY

Some of the leading visionaries of Internet architecture are among the enthusiastic proponents of metadata. Most prominently, Tim Berners-Lee, credited with inventing the World Wide Web,¹⁰¹ articulated his vision for a metadata-enabled “Semantic Web” in 2001:

Most of the Web’s content today is designed for humans to read, not for computer programs to manipulate meaningfully. . . . The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users.¹⁰²

Berners-Lee and the World Wide Web Consortium (“W3C”), which he leads, continue to advocate for the development and adoption of metadata standards required to realize this vision.¹⁰³ But they have met some skepticism, much of it focused on the question of metadata accuracy.¹⁰⁴ Metadata that is perfectly compliant with standards that ensure its interoperability and durability may nonetheless be worse than useless if it is not accurate. To take a simple example: a descriptive tag that identifies a photo of a cat as a photo of a dog will be annoying—not useful—to subsequent collectors of dog photographs. The issue of accuracy is of course even more serious when metadata relates to legal relationships.

How would a would-be re-user of a digital photograph know whether the metadata specifying copyright notice and license terms were embedded by someone who is or has the authority of the copyright holder? Or if, instead, the metadata misidentified the copyright owner and misrepresented the terms under which the true owner was willing to license the work? The short answer is that, given the current state of copyright law in the United States

101. See, e.g., Tim Berners Lee (1955-), BBC HISTORY, http://www.bbc.co.uk/history/historic_figures/berners_lee_tim.shtml (last visited Feb. 6, 2013).

102. Tim Berners-Lee et al., *The Semantic Web*, SCIENTIFIC AMERICAN (May 17, 2001), available at <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>; *Semantic Web Frequently Asked Questions*, W3C (2009), <http://www.w3.org/2001/sw/SW-FAQ>.

103. *Semantic Web*, W3C, <http://www.w3.org/standards/semanticweb/> (last visited Sept. 9, 2012).

104. For an especially pithy example, see Doctorow, *supra* note 62. For an overview of metadata quality issues and existing literature, see Jung-Ran Park, *Metadata Quality in Digital Repositories: A Survey of the Current State of the Art*, 47 CATALOGING & CLASSIFICATION QUARTERLY 213, 225 (2009) (concluding that “the rapidly growing body of digital repositories calls for further investigation of metadata quality” and that “[f]uture studies lie in identifying factors behind inaccurate, inconsistent, and incomplete metadata creation and application”).

and elsewhere, a user cannot answer these questions on the basis of metadata alone. The question of copyright ownership can be complicated and context-specific, depending on facts about the creative process (was the photo taken by the owner of the camera, or by a friend?); about preexisting relationships (was the photographer being employed to take the photo?); and about subsequent events (did he assign his copyright to someone else?).

These problems should give us pause as we consider the feasibility of metadata-based solutions to the problems caused by atomistic copyright. But the problems are not new. The same types of misrepresentations and uncertainties can produce a faulty printed copyright notice in a paper book, or even in the records of the Copyright Office. Since copyrights arise with the fixation of an original work of authorship—and not when notice is provided or when the Copyright Office says so—genuine copyright interests can quite easily coexist with false notices and registrations.¹⁰⁵

Although the basic structure of the Copyright Act contributes to the possibility of conflicts between true copyright ownership and official-seeming representations about ownership, the Act is not entirely inattentive to the problem. If the person who provided the false copyright notice had fraudulent intent, he has committed a crime under U.S. copyright law.¹⁰⁶ Making a knowingly false representation in an application to the Copyright Office is also a crime.¹⁰⁷ The provisions regarding “Copyright Management Information,” added by the DMCA, further prohibit knowingly providing false copyright management information with intent to “induce, enable, facilitate, or conceal infringement.”¹⁰⁸ It seems unlikely, however, that the threat of legal sanction will fully deter (or punish) false provision of copyright notice—especially in the form of digital metadata often supplied

105. See 17 U.S.C. § 411(b)(1) (2010). Knowingly submitting a materially inaccurate application will render the resulting certificate of registration insufficient for purposes of instigating an infringement action, however. *Id.*

106. 17 U.S.C. § 506(c) (2010) (“Any person who, with fraudulent intent, places on any article a notice of copyright or words of the same purport that such person knows to be false, or who, with fraudulent intent, publicly distributes or imports for public distribution any article bearing such notice or words that such person knows to be false, shall be fined not more than \$2,500.”).

107. 17 U.S.C. § 506(e) (“Any person who knowingly makes a false representation of a material fact in the application for copyright registration . . . , or in any written statement filed in connection with the application, shall be fined not more than \$2,500.”).

108. 17 U.S.C. § 1202(a) (2010) (“No person shall knowingly and with the intent to induce, enable, facilitate, or conceal infringement (1) provide copyright management information that is false, or (2) distribute or import for distribution copyright management information that is false.”).

anonymously and with inaccuracies that may result more from carelessness or legal ignorance rather than from fraudulent intent.

Some metadata advocates have proposed technical solutions that address this problem not by directly guaranteeing accuracy or punishing inaccuracy, but rather by facilitating the provision of contextual information that can help users assess the accuracy of metadata and also increase the accountability of metadata providers. For example, in a paper describing the Creative Commons Rights Expression Language (“ccREL”)—an XMP-compatible standard for machine-readable rights-related metadata—Creative Commons identifies the issue of accountability and its special relevance where metadata is embedded in photos and other types of “free-floating content.”¹⁰⁹ In this context, “the method for expressing metadata should facilitate providing publisher accountability at least as strong as the accountability of a Web page with a well-defined host and owner.”¹¹⁰ How would this work? Here is the CC approach:

We handle accountability for free-floating content by connecting any free-floating document to a Web page, and placing the ccREL information on that Web page. Thus, publishers of free-floating content are just as accountable as publishers of Web-based content: rights are always expressed on a Web page. The connection between the Web page and the binary file it describes is achieved using a cryptographic hash, i.e. a fingerprint, of the file. For example, the PDF file of Lawrence Lessig’s “Code v2” will contain a reference to <http://codev2.cc/download+remix>, which itself will contain a reference to the SHA1 hash of the PDF file. The owner of the URL <http://codev2.cc/download+remix> is thus taking responsibility for the ccREL statements it makes about the file.¹¹¹

In short, CC recommends that rights-related metadata include not merely assertions about the copyright owner and license terms applicable to a photo or other piece of content, but also an association with a webpage that can in turn provide a would-be user of the content with some indications of the identity and reliability of the provider of the metadata. This solution is not perfect (as Internet domain name holders can also be difficult to verify and locate), but it does seem to make the accuracy problems associated with digital rights-related metadata akin to the pre-existing threat of false copyright notice on non-digital works. Printed copyright notices and licenses

109. HAL ABELSON, ET AL., CCREL: THE CREATIVE COMMONS RIGHTS EXPRESSION LANGUAGE 11 (Creative Commons 2008), <http://wiki.creativecommons.org/images/d/d6/Ccrel-1.0.pdf>.

110. *Id.*

111. *Id.* at 17.

are not useless in general—and neither is copyrighted-related metadata—merely because it is conceivable that they could be falsely provided by an impossible-to-locate publisher who cannot be held accountable for the falsehood.

In sum, the problem of inaccurate rights-related metadata (like the problem of inaccurate metadata more generally) is a real concern that has rightly drawn the attention of both the advocates and critics of increased use of metadata to manage digital files. But this concern should be evaluated in light of the alternative: files distributed with no copyright information attached to them whatsoever, in a legal environment in which they are automatically protected by copyright. Even if metadata merely give would-be re-users some clues as to the identity of the copyright owner and her preferences—which users may want to confirm through further investigation facilitated by the recommended practice of providing web-based as well as embedded information—in many cases that will be much more information than users would otherwise have had.

Eventually, the accuracy challenge may best be addressed by combining copyright-related metadata with the types of revamped registries that have been proposed by copyright reformers and explored by entities like PLUS and Creative Commons.¹¹² Metadata could both provide a copyright notice and license terms and refer to a centralized registry—perhaps one maintained by a governmental or private body that makes some attempt to verify the ownership claims (akin to the functions provided by land records offices and private title insurance companies in the real estate context).¹¹³

VI. CONCLUSION

This case study of the current practices and prospects for embedding durable copyright-related metadata in digital photographs suggests both the potential and challenges for this type of technological solution to the problems posed by atomism in today's digital copyright environment. We can now imagine how, if the emerging technical and legal standards were combined in a consistent and ubiquitous way, photos that would otherwise be orphaned could be used and reused in compliance with the wishes of their copyright owners—but without costly searches and negotiations. Difficult challenges remain, however, the most intractable of which may be ensuring

112. See *PLUS Registry*, PLUS COALITION, INC., <https://www.plusregistry.org/> (last visited Sept. 9, 2012); Eric Steuer, *Creative Commons Explores a Digital Copyright Registry System*, CREATIVE COMMONS (May 22, 2008), <http://creativecommons.org/press-releases/entry/8306>.

113. See Riecks et al., *supra* note 17, at 3.

metadata accuracy. But this problem is nothing new—either to the metadata community or to copyright law. Indeed, the best way forward may be to combine improving practices for the use and handling of metadata with proposals for revamping copyright registration—using the best of both decentralized and centralized information to provide notice that is interoperable, persistent, comprehensible, and accurate. Such a system could mitigate the costs of atomistic copyright without unduly compromising authorial autonomy and control. Removing the remaining barriers to realizing this vision may be costly and complicated, however. Existing platforms and the software that powers them do not yet incorporate all of the functionality that I have described. Addressing atomism with automation is an attractive vision; but the difficulty and expense of solving copyright challenges with technology should also figure into debates about more fundamental copyright policy reform.¹¹⁴

114. Many of the articles collected in this issue include reform proposals relevant to the problems addressed here, as do, for example, Samuelson et al., *supra* note 2; and Sprigman, *supra* note 8. Most dramatically, perhaps, Justin Hughes has challenged us to rethink the assumption that all photographs in this digital and atomistic age are automatically copyrightable. He suggests that “[a]s we move toward the ubiquitously recorded world in which we are not only constantly photographed, but we ourselves continuously record photographic images and video streams, the whole idea of originality in photography may come under renewed scrutiny.” Hughes, *supra* note 20, at 413.

