

Metadata Persistence Not Yet Part of Photoshop CS3's "Save for Web..."

By Ron Roszkiewicz

The importance of copyrights cannot be over-emphasized in today's atmosphere of casual digital piracy. Everyone who creates digital content for the Web knows the relative ease with which digital assets can be copied and reused. Increasingly, copyright information is being embedded in files, giving owners potential recourse when those files end up somewhere they shouldn't. The expectation is that each time such metadata is added to a file it will live on, regardless of how it might be edited or how many times it may be copied – a condition known as persistence.

We expect vendors like Adobe to support persistence for copyright information in files created with their applications. This expectation will grow as more users acquire the discipline it takes to include metadata, build templates and process files efficiently. In general, it looks like Adobe is on the right track with this ongoing support. With the release of CS3, the company is adapting to the needs of the Web designer and mobile content provider – with new integrated tools such as Device Central. Unfortunately, all is not as it seems in XMP-Ville.

In both Adobe Illustrator CS3 and Photoshop CS3 there is a File menu item called "Save for Web or Devices..." (Before Device Central, this was simply called "Save for Web...") The command brings up an interactive window for tweaking the file to match the desired quality requirements, at a file size that can be lived with. With CS3, this has been further enhanced by a built-in link to Device Central. Clicking on Device Central launches the application, emulates cell and smart phone screens and transforms the image to the target screen size. The problem is that using this process results in only Keywords and one Copyright field's contents being saved. Other metadata is lost. It was rumored that this was done in order to make the file as small as possible. However, XMP xml packets weigh in at only 4K – making such concerns increasingly ludicrous.

In CS3, Adobe added an additional option to the "Save for Web or Devices..." process. After clicking the Save button, a new dialog appears prompting the user for a

file name for the new downsampled file. At the bottom of the screen is a Settings pop-up with more options, including "Other." Choosing this brings up another dialog with, among other things, a checkbox for "Include XMP" – whose settings can be saved for repeated use. Unfortunately, this option only yielded some but not all of the filled-in metadata. The copyright symbol and the year was missing but, curiously, not the name of the copyright holder. The data in Keywords, Copyright Choice and Copyright Info URL were also missing.

The bottom line is that users who really want to standardize on metadata for their assets must resort to a more complex workflow (downsampling and using the "Save As..." command) to create images for the Web. Some techno-savvy users have created command line approaches to stuffing the metadata back into the file on the other side of the transformation but such an approach is often beyond the reach of typical users.

One of the main reasons given for the casual attitude regarding persistence is that there is no hacker-proof way to preserve the integrity of metadata in files. There are many ways to defeat the system, including

manually changing the metadata by opening the file and then re-saving it. That may be true. It may be that the only way to truly copyright a digital file is with an integral watermark. However, the argument dismisses the fact that without metadata persistence internal DAM and CMS workflows themselves are broken. Metadata without encryption may just be a mild deterrent for pirating a free roaming digital asset. However, within a company's workflow, on either side of the firewall, consistency and persistence for many users is what makes the case for metadata in the first place.

The fact that Adobe is standardizing on the look-and-feel of application interfaces throughout the Creative Suite is laudable. PDF itself is becoming increasingly useful inside Creative Suite applications, as it has become an effective outside exchange format. It is also wonderful that Adobe is encouraging metadata use by giving users more and better ways to define and embed it. We are willing to accept the fact that Adobe is wrestling with how to protect and secure data in files that are human-readable and built on open standards. However, the persistence of "Save for Web..." metadata flaw is an embarrassment – which we hope will be remedied without delay. It breaks digital asset management and repurposing strategies by altering what metadata survives when derivative files are created. It is time to move this topic out of the forums and onto the "Things to Fix" priority list. **TSR**

PRISM 2.0 Specification Released for Public Comment

By Linda Burman

On July 12, the IDEAlliance PRISM (Publishing Requirements for Industry Standard Metadata) Working Group released the final draft of the Version 2.0 specification for public comment. The draft is available from the PRISM [Web site](#), and public comments may be submitted through September 15, 2007. The specification will be finalized at the September PRISM meeting.

This major revision addresses new requirements from publishers and other media companies to deliver content online (Web and mobile), to provide a mechanism to encode multimedia objects with

PRISM XMP, and to enhance the PRISM controlled vocabularies to provide more specific metadata based on delivery media type, presentation format and genre.

The PRISM specification defines a set of metadata vocabularies for managing, aggregating, post-processing, tracking and multi-purposing magazine, news, catalog and mainstream journal content. In addition PRISM provides a framework for interchange, interoperability and preservation of content and metadata.

The PRISM working group was founded in 1999 by a group of publishers, content providers, system integrators, software developers and content aggregators

under the umbrella of the GCA (Graphics Communications Association, now IDEAlliance). It addresses the need for a metadata specification that provides the appropriate “tags” for complex, multi-part articles such as those found in magazine and journal content.

The founding members faced challenges such as re-use of content in multiple media types, rights and contract management, better access to content archives, and faster, less expensive exchange and integration of disparate sets of content across the enterprise and with outside business partners. They believed that developing and adopting a standard set of XML metadata would assist them in managing and automating their labor-intensive content workflow processes.

The Business Case

Publishers see the effective use of metadata as the key to cutting costs from production operations and growing revenues by leveraging new electronic distribution methods. The PRISM Specification represents agreement across the publishing community on the importance of a standard XML metadata vocabulary.

Metadata is descriptive information that is machine-readable data added to human-readable content to simplify the automation of routine, but crucial tasks, such as content searching, determining rights ownership and personalization. Providing these capabilities has also captured the interest of organizations outside the publishing domain that are dealing with the issues of creating, managing and distributing large amounts of content.

“Our business is vitally dependent on customers being able to find what they need by searching the metadata, including keywords,” said Dr. Lisa Frumkes, Linguistic Analyst at Getty Images. “PRISM’s extensive subject description capabilities, plus its handling of basic rights and permissions, help support our long-term metadata needs.”

Why PRISM 2.0?

Although the current version (1.3) addressed many industry requirements, several major publishers have been actively engaged in making the standard more reflective of new media realities. Time Inc.’s E-Content Management Group found considerable value in using the current version the last 3 years, but has invested considerable time helping develop PRISM 2.0. Dai Ichikawa, the group’s Digital Asset Systems Manager,

commented that this was in part due to the growth of content on the Web and mobile devices. “This content has unique properties and metadata needs that PRISM currently does not address,” he said.

Ichikawa also described several of the changes in 2.0: “Several new metadata elements are being added. For example, the *prism:postdate* element would capture the date an article was posted on a Website where issue dates, as used in print magazines, do not exist. One of the main goals in updating PRISM is to make it more adaptable and flexible to accommodate a larger variety of content and business needs while keeping usage guidelines well-defined. To this end, an element such as *prism:genre*, with an extensive controlled vocabulary, can be used to describe a wide range of content — abstracts, blogs, cover stories, feature packages, product descriptions, etc.”

Stephanie Salmon at U.S. *News & World Report* had similar reasons for working towards the 2.0 version. “The driving factors for our interest in upgrading PRISM include the need to better reflect our publishing platform expansion,” she said. “U.S. *News* is published

in a variety of ways including print, Web and mobile and it’s important to us that the metadata standard be compatible with those platforms. We also felt the controlled vocabularies needed to be expanded to include types of content that are were not addressed in the current version, such as blogs.”

Salmon continued, “Some of the changes [in 2.0] that will be particularly useful for us address Web content. For example the element *prism:alternateTitle* will allow us to capture a headline that has been altered from the original to suit a different platform. Since the same content may be sent out through a variety of platforms, each necessitating a different headline, this element allows us to capture that metadata.”

Lee Vetter, Publishing Technologies Specialist at McGraw-Hill, also described the driving factors behind his company’s involvement in 2.0. “We need to be able to have one metadata source that has the flexibility to deliver to a variety of platforms in a speedy manner,” he said. “With PRISM 2.0, we will be able to deliver to print, Web and mobile [platforms] without creating new source files.” **TSR**

Educational Publishing Field Narrows

By Anita Malnig

On July 16, Houghton Mifflin Company announced the acquisition of Harcourt Education from Reed Elsevier for \$4 billion — \$3.7 billion in cash and \$300 million in common stock of Houghton Mifflin in Riverdeep Group PLC (Houghton Mifflin’s parent company). The merger narrows an already streamlined field. The other major players include Wiley & Sons, McGraw-Hill, Pearson and Scholastic. Harcourt imprints include Harcourt Achieve whose recent textbook Saxon Math Course 1,2,3, won a distinguished achievement award from the Association of Educational Publishers. Other imprints include Harcourt School Publishers; Holt, Rinehart & Winston; and Harcourt Trade Publishers.

Among the changing business models within the sphere of educational publishing is a movement away from the all-encompassing textbook to smaller texts with numerous Internet-based ancillary materials. With this trend, continued consolidation may prove to be increasingly effective for publishers’ bottom lines. According to Barry O’Callaghan, principal shareholder of Houghton Mifflin, “Together Houghton Mifflin and Harcourt

will provide a more customized, diverse and innovative range of products to meet the changing need of educators and students.”

Tony Lucki, chairman, president and CEO of Houghton Mifflin, and former CEO of Harcourt Education and Harcourt Inc., will head the combined company, which represents over \$3 billion in revenue. He promised “more customized and innovative content,” but gave no specific indications as to the company’s strategic direction.

Can fewer companies provide more? Both companies are respected as traditional publishers, which does not necessarily translate into successful innovation. According to Sir Crispin David, CEO of Reed Elsevier, the sale price “recognizes the exceptional quality of the Harcourt assets and of its management and staff.” Whether that means success in the face of a rapidly changing market, only time will tell. **TSR**

Anita Malnig is currently researching trends in the educational publishing market for a future article in *THE SEYBOLD REPORT*. She can be reached at amalnig@sbcglobal.net.