Making the Case for Embedded Metadata


Kari R. Smith
Massachusetts Institute of Technology


Ulla Bøgvad Kejser
Kongelige Bibliotek

Making the Case for Embedded Metadata in Digital Images

Kari R. Smith, Massachusetts Institute of Technology, Cambridge, MA, USA; Sarah Saunders, Electric Lane, London, UK; Ulla Bøgvad Kejser, The Royal Library, Copenhagen, Denmark.

Abstract

This paper discusses the standards, methods, use cases, and opportunities for using embedded metadata in digital images. In this paper we explain the past and current work engaged with developing specifications, standards for embedding metadata of different types, and the practicalities of data exchange in heritage institutions and the culture sector. Our examples and findings support the case for embedded metadata in digital images and the opportunities for such use more broadly in non-heritage sectors as well. We encourage the adoption of embedded metadata by digital image content creators and curators as well as those developing software and hardware that support the creation or re-use of digital images. We conclude that the usability of born digital images as well as physical objects that are digitized can be extended and the files preserved more readily with embedded metadata.

Introduction

Over the last two decades, archives, libraries, museums and other cultural heritage institutions have increasingly opened up their visual and text-based physical collections to users by making digital images of the works available online. Likewise, these institutions are making their collections of born digital images available. To facilitate users' search and discovery of these objects, users are supplied with contextual information – metadata – describing aspects of the artistic, historic, and documentary value of the work and/or the image and the rights associated with use.

Today digital images are intensively used, exchanged and shared over the Internet – on social media and networking sites, such as Facebook, Pinterest, Picasa – re-used and crowd sourced by other organizations and users, and archived for use at some future date. There is also growing research interest in aggregating and processing image data and metadata from different sources to create new knowledge. To support this trend, institutions are exploring ways of making the images and their metadata available to users.

Heritage institutions typically store descriptive, administrative and rights data about a work in collections management databases and digital asset management systems (DAMS). These systems also hold metadata needed for managing, preserving and using the digital assets. Such systems provide effective ways of managing the images online. However, when the images are downloaded they lose the connection to the stored metadata. To meet this challenge institutions are investigating how to better make use of the possibility of embedding metadata directly in the file.

Approach and Methods of Inquiry

Our work in the areas of developing and using embedded metadata led to several use cases that we describe here as our approach and methods of inquiry. This included to a) review the current use and value of embedded metadata in the heritage sector and detailing use cases b) investigate the issues raised in practice when embedding metadata in heritage images and discussing ways of working around inconsistencies and c) discuss solutions to enable and promote the use of embedded metadata in the future. In this paper we examine cultural heritage institutions' use of embedded metadata and summarize current challenges and best practices. We describe current developments in metadata schemas for embedded data and indicate how these can benefit the cultural community in the future.

Embedded metadata

Digital media files contain a great deal of information about themselves - they have to in order to be useful. Without such information as file type, compression algorithm, and color profile, viewing tools would be unable to decode and display an image. Embedded metadata is not limited to technical instructions - a description of the content can also travel within the file. Properly applied, embedded descriptive metadata can be as easily understood and used as technical metadata. Knowing who created the object(s) shown in a digital image can be as easy as knowing when that image file was created.

Metadata is often categorized according to the function of the metadata: descriptive, technical, structural, administrative, rights, and preservation. When describing cultural objects - works - data can broadly be categorized into the following headings; administrative, descriptive and rights data. Moreover, these data types can apply separately to objects and images of the objects (though in areas such as 'content description' the data may be the same).

Metadata is arranged in schemas that consist of elements and their associated values. A metadata schema standardizes the names of the elements and their meaning, and it may also set the rules for formatting the values and indicate the use of vocabularies. For example, an element can have the name "Date" with the meaning "the date the image was captured" and specify that the value must be a standardized format, such as "DD:MM:YYYY hh.mm.ss".

The use of standardized metadata schemas facilitates exchange of metadata across different software and systems (compatibility). If two systems support the same metadata schema then the metadata elements and values formatted by the first system can be read, translated (mapped), and written directly to the succeeding system. If the two systems support different schemas then a more advanced mapping is required.

Many common image formats such as JPEG, TIFF, JPEG2000, PNG, and also the document format PDF, allow for embedded metadata, having specific segments of their file structure set aside for metadata. In a TIFF file, for example, this information is stored in the image file directory (IFD) [1] in distinct byte groupings called tags. Technical metadata such as, camera settings, time and date, image size, compression, name of camera, and color information are added to the Exif tag by devices at the time an image file is created. In some cases, such as RAW image files, software will write metadata to separate (sidecar) files which can be embedded before the images are shared. Keeping metadata in a sidecar file works well if the user of the file has appropriate software to decode the sidecar metadata file. For file exchange or distribution embedding the metadata directly within the source image file is a better option.

Standards, schemas and formats for embedded metadata

Metadata can be embedded in image files using Adobe's XMP (Extensible Metadata Platform). XMP, an open source RDF/XML based ISO standard (ISO 16684-1:2012), is a labeling technology and a framework (container) that uses existing metadata schemas to embed data inside the image file. The extensible nature of the framework means that custom schema can be carried as well as the standard schemas supported by Adobe for use in its own software. XMP was developed to secure interoperability and is now widely supported in most file browsers, media software and web services. XMP allows multiple descriptions of a resource in various languages through the use of ISO 639-1 country codes in the xml:lang attribute.

The Exchangeable Image File (Exif) [2] is a non-extensible standard agreed by camera and scanner manufacturers for data created by the camera or scanner at the point of image capture; it includes technical details such as date and time of capture, pixel dimensions and camera settings, and in some cases the name of the photographer and the GPS coordinates. Exif and XMP data coexist in the image file and sometimes Exif data is picked up in XMP.

Over the years more metadata schemas for embedded descriptive metadata have been developed to serve different domains and purposes. Today the most commonly used schema is IPTC (International Press Telecommunications Council) [3] Core and Extended. Initially developed for the press industry, IPTC has been adopted by many different communities and software developers.

In the cultural heritage sector, the Dublin Core Metadata Initiative (DCMI) [4] is also widely used. Dublin Core metadata was originally created to describe digital assets on the Web and so there is little ability to distinguish describing the asset (digital file) from the content in the digital file, such as the image. There are tools that enable automatic synchronization between selected IPTC Core fields and DC elements, though these mapping can be imprecise. An Example of this is Dublin Core Title, which is often used for 'Title of an Artwork', and mapped to IPTC Core Title field, although the IPTC schema has a specified IPTC Extension field 'Title (Artwork or Object). VRA Core 4.0 XMP info panel was developed to enable the transfer of granular embedded data in heritage images. Information about the work (e.g. artist, title of artwork) and the image of the work (e.g. date of image capture, photographer) are held in separate fields. It is customizable to specific needs, e.g. with Adobe Creative Suite tools; offers data import and export functions; and synchronizes with key IPTC and Dublin Core fields. It was developed for use in Adobe Bridge.

Tools for handling embedded metadata

If a file is copied or edited, its technical metadata may be updated automatically by the software being used. Descriptive and administrative metadata, such as IPTC is added by the user and requires software capable of parsing the various metadata formats. The most basic tools with this ability are system file browsers, such as: Mac Finder and Spotlight, and Windows File Explorer, which also allow metadata to be edited and searched. At consumer-level applications allow for editing of basic metadata, such as Title and Keywords, while more professional image processing and management software allow editing of the full IPTC schema. Adobe Photoshop and Bridge, support many published schema and provide Software Developer Kits (SDKs) for creating custom schema and editing forms. Several open source tools exist to fully analyze and export embedded metadata.

Understanding how applications interpret and display embedded metadata schema is important for establishing best practices for metadata creation and implementation. To maintain backward compatibility, software usually duplicates the same data in matching properties in Exif, IPTC IIM (legacy), and XMP. For tools that analyze embedded metadata see [18].

History of IPTC photo metadata standard

The IPTC (International Press Telecommunications Council) is the main standards body for data embedded in the image file. The Information Interchange Model (IIM) schema was created in the early 1990's for the news industry. Since the creation of Adobe's XMP framework in 2001 the IPTC schema has been designed for use in XMP. IPTC Core was created in 2004, followed by IPTC Core 1.1 and IPTC Extension 1.1 in 2008.

The use of IPTC has extended beyond the news industry to include photography, image libraries, print and web publishing and cultural heritage. IPTC is supported by Adobe and other software products and is widely used in photography and image library workflows and in cultural heritage.

IPTC Extension created additional fields to describe artworks and objects in the image, removing some of the ambiguity in the IPTC Core. Artwork or object fields for title, description, creator and date of creation of an artwork or object were added. The IPTC Extension fields are not yet widely used in the heritage sector, partly because data about the artwork or object is handled in the first instance by curatorial and cataloguing staff who are not image specialists. Many institutions are currently engaged in an effort to coordinate image and data workflows for photography, collections management and DAMS, and the understanding of the image workflow with its associated data is increasing.

Recent proposed changes to the law on Orphan Works in Europe and the US have increased pressure on rights holders to protect their work using embedded metadata. However, metadata stripping is a problem particularly in the web environment, and IPTC has set up EMM (Embedded Metadata Manifesto) [5] to encourage retention of metadata. Metadata stripping has also been addressed by Metadata Working Group (MWG) [6] representing major hardware and software companies, and at the UK Government's Copyright Hub.

Since the launch of IPTC Extension some gaps relating to heritage objects have become apparent. An increased interest in embedded data among practitioners in the sector has led to a proposal to add new fields such as a text formatted ‘circa date’ field, and a web link to authoritative collections management data.

The IPTC has agreed criteria for adding new heritage fields to IPTC Extension; they should convey rights and descriptive information helpful to the non-specialist user of the image, but would not contain deep collections management data. They may be considered as display fields, providing useful labels for general use.

The VRA custom XMP info panel, an Adobe Photoshop and Bridge extension for artwork description, struck a similar balance between granularity and practicality by only using display properties from the VRA Core 4 schema and only supporting the description of one artwork. While this approach met the needs of most users, others have expressed the need to describe multiple artworks and to record attributes like Creator Attribution, Role and Vocabulary.

The IPTC/SCREM Project

The project SCREM (SChema for Rich Embedded Metadata for Heritage Media Files) has emerged from needs identified by the IPTC and the VRA info panel project. Where IPTC and VRA have focused on users within their communities, SCREM seeks input from a wide variety of heritage institutions from around the world. The goal is to reach a general consensus on an embedded metadata standard that serves the needs of a wide range of users, avoiding the proliferation of incompatible local solutions. The IPTC/SCREM project will roll out in two stages by first identifying candidate fields for IPTC Extension (by June 2014) and second by building upon the new IPTC Extension properties by creating a more granular schema (SCREM) for use between and within heritage institutions. A custom XMP panel would be created (timetable to be decided) [7] which contains fields drawn from other schemas and will be designed for maximum interoperability.

Practitioners in heritage organizations have been consulted and the IPTC candidate fields are under review by the SCREM group. Use cases identified for IPTC heritage fields by participants include: documentation for images (on removable media or downloaded from the internet) for publishers, designers, researchers, students, general public, education facilities; and data upload to aggregate sites such as Europeana.

Use cases for the wider SCREM schema include: exchange of data between heritage organization departments; helping streamline image production; allowing photographers to enter data; reducing wear and tear on vulnerable objects; bridging gaps in organizations where the collections management system is not accessible; providing data for researchers and art historians who manage their own collections of images and data; encouraging organizations to share metadata, held by small organizations working without collections management systems who store data on spreadsheets.

Digital preservation and digital archiving

Preservation and long-term use of digital images relies not only on the technical but also on the intellectual knowledge of the files. Overtime, digital images are at risk of falling out of curatorial care. This might be for just a year but even a brief lapse in curation can result in often the digital images being separated from the descriptive and administrative metadata. Images may be single items or linked ones such as pages in a book or images in a sequence, and the metadata about them is critical to making the digital images useful and useable in the future -- the goal of preservation.

Embedding metadata into the digital files can be essential to helping a downstream user or archivist to understand the content, the creator and the usages rights of the files. The Metadata Working Group (MWG) and the Picture Licensing Universal System (PLUS) coalition both state the long-term preservation and access to digital images as part of the reason for their work and cooperation. The PLUS coalition [8] explicitly discusses the need to use embedded metadata to manage and preserve digital images. PLUS has standardized rights statements with PLUS ids that can be embedded using XMP. The Metadata Working Group’s expressed goal to preserve metadata that in the past has gotten lost in the translation to camera software specifications also highlights the role that industry is having in this discussion.

Use Cases and Inquiry

Use Cases at The Royal Library in Denmark (KBDK)

To illustrate some of the practical implications of using embedded metadata in image production and management workflows we describe an effort carried out at the KBDK to investigate the usefulness of embedded metadata for two purposes.

The first case deals with the library wants to embed a subset of the descriptive metadata in the files that will follow the images when they are downloaded and displayed by common user applications in order to enhance the usability of the image files.. These metadata include: dc:Title, dc:Creator, dc:Date (iptc:Date Created), dc:Description, dc:Subject (iptc:Keywords), and dc:Rights (iptc:Copyright Notice).

Since many of the images originate from digitization projects there was a focus on ways of embedding metadata that would differentiate data about the ‘work’ from data about the ‘image of the work’. Some fields contain shared data; the description of the content of the image (e.g. of a painting) will be the same as the description of the object shown. If an image depicts several works such as sculptures, descriptions will differ. When an image of an object is digitized, there exists a creator for the digital image, and a creator for the work. Incorrect usage of the creator fields can lead to confusion when automatic labels are applied, which some applications do. ‘Photo taken by Leonardo Da Vinci’ is not a convincing label.

IPTC Core and Dublin Core were not designed to express the duality that occurs for images of objects. However, IPTC Extension includes fields for describing the Artwork-Object in the image, and likewise, VRA Core has separate fields for information about the work and the image of the work, as well as fields for administrative metadata. So to achieve the required granularity of the metadata KBDK would use IPTC Extension or VRA Core. However, IPTC Extension and VRA Core are only rendered by more advanced software, such as Adobe CS tools. Only IPTC Core is implemented by widely used applications.

Looking at this dilemma between required granularity and renderability of the metadata schema, renderability was considered the most important criteria and the Library therefore decided to use

IPTC Core. As explained, the IPTC Core standard describes the image of the work - not the work itself. However, KBDK does not consider digitization, when the aim is to produce faithful reproductions of a work, to change the intellectual content of the work. Because the digital reproductions are not recognized as new works it was therefore decided to bend the standard in the sense that the IPTC Core metadata would describe the work and not the image of the work, thus appropriating the schema fields for our own custom use.

While this implementation could work for internal purposes, in practice it raises a number of problems. The IPTC Core date formatted field does not allow for circa dates expressed as text format e.g.. 1960-67 or ca 1800. Placing the object created date in IPTC Core would not only be inconsistent in this case, but also could cause problems on exchange with other systems, e.g. if the date 1934 is transformed into a specific date/time format such as 1934-12-01 14:25:20.

The library therefore decided to use the standard as intended and then assemble by concatenating all the metadata relating to the work, including the name of the creator, and the date the work was created in the Description field (see table 1). This is a practice employed by many image libraries in the cultural sector, and has the advantage of displaying a reasonably full set of data in many systems which do not display up to date IPTC Core and Extension fields. This practice also leaves open the addition of granular data at a later date. For accreditation purposes, the creator of the image was set to be ‘The Royal Library, DK’.

In the second use case at the KBDK the goal was to gain efficiencies in the image production workflow by enabling repro-photographers to embed certain descriptive metadata during the digitization process. Currently, technical metadata is embedded via camera and scanner systems, and image processing software. The image files are then delivered to the library’s cataloging section where descriptive metadata records are added through a digital asset management system (Canto Cumulus). This system is capable of importing a range of standardized embedded metadata schemas, such as Exif and IPTC Core, directly and through mapping also custom schemas, including the VRA Core. Likewise, the system can export metadata from the records to the image files.

To assure the quality of the metadata, normally only librarians and cataloguers provide descriptive metadata to the images. However, repro-photographers can be authorized to add some types of metadata that describe the physical nature of the work - whenever possible from predefined vocabularies - such as the type of material, technique, size/dimensions, and any inscriptions (see image 4). The reason for this delegation is that otherwise, not only the repro-photographers, but also the cataloguers would need to handle the object that was digitized, which would not only increase production time, but also the wear and tear of the collections. The repro-photographers are also responsible for embedding metadata about the applied imaging quality. Currently, KBDK is working to insert a URI that points to a database in which image quality analysis reports for the related digitization equipment and time are stored. The VRA Core is used to capture these types of detailed metadata.

Other use cases in Heritage Sector

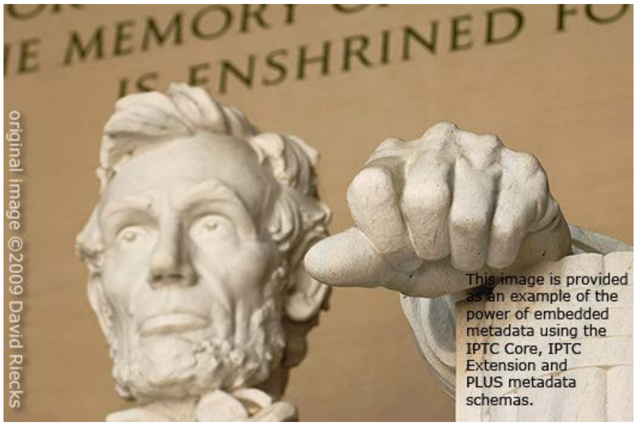
In addition to the way that KBDK is using embedded metadata, there are many other ways that embedded metadata is being used and implemented in the Heritage Sector generally. When an arts organization reorganizes its images prior to purchasing a new collections management system and DAMS, for example, data can be embedded in the images to assist users in identifying and discovering images using a simple image browsing software. In addition, picture research can be conducted more efficiently when a picture researcher downloads low resolution images from a number of different art libraries for use in a publishing project. The descriptive data and source of the image are embedded in the images, which are viewed in image viewing software. For new digital photography, the production staff can embed key information about an object in the image and this is ingested into the collections management system and checked by curators.

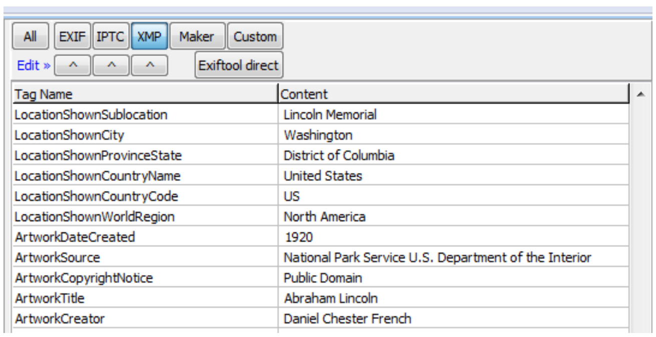
Use cases for the wider SCREM schema (which contains fields drawn from other schemas and will be designed for maximum interoperability) include: exchange of data between heritage organization departments; helping streamline image production; allowing photographers to enter data; reducing wear and tear on vulnerable objects; bridging gaps in organizations where the collections management system is not accessible; providing data for researchers and art historians who manage their own collections of images and data; encouraging organizations to share metadata, held by small organizations working without collections management systems who store data on spreadsheets.

An example of an innovation using the power of embedded metadata is the MetaShotPpt [9]. PowerPoint is widely used to create instructional presentations featuring images of cultural heritage works. The process of inserting images and manually entering descriptive information can be quite time consuming. MetaShotPpt is an open source tool which streamlines this process by taking a folder of image files and automatically importing each image into a single slide, and if the file has descriptive embedded metadata, copying it into the presenter notes. Enhancements to MetaShotPpt will allow metadata to be added as a caption adjacent to the image.

Digital Forensics and Preservation

Digital forensic techniques and tools are increasingly being used in the heritage and business sectors to recover digital image files that were placed on media no longer in use. The main technique used is to create a disk image, a process that copies the contents of the media at a bit level to a standard format. Without needing to open the files in the disk image, many features of the content can be discovered, including metadata that is embedded in the files. Using the BitCurator Tools and pyExifTool we ran the following sample image and were able to see the metadata without needing to open the image files (see Figure 1 and 2). This is very useful to a curator or archivist to ascertain what files and content is on a media carrier or a computer hard drive. Through testing and evaluation of tools and workflows of digital forensics and preservation, we analyzed how embedded metadata might extend the usability of digital objects and make work processes more efficient by negating the need to open digital images files before appraisal [10].





Tag Name	Content
LocationShownSublocation	Lincoln Memorial
LocationShownCity	Washington
LocationShownProvinceState	District of Columbia
LocationShownCountryName	United States
LocationShownCountryCode	US
LocationShownWorldRegion	North America
ArtworkDateCreated	1920
ArtworkSource	National Park Service U.S. Department of the Interior
ArtworkCopyrightNotice	Public Domain
ArtworkTitle	Abraham Lincoln
ArtworkCreator	Daniel Chester French

Figure 1*: IPTC Core/Extension with artwork/object metadata.*
Figure 2*: ExifToolGUI display (detail) of XMP metadata from Image.*

Discussion

The exchange of metadata can be complicated by different factors. In some cases schemas use different property names for the same concepts (subject, keywords, tags;) creator, artist and software use different labels for the schema properties. For example XMP dc:subject might be labeled "Keywords", "Subject", or "Tags" and dc:description might be labeled "Subject", "Description", or "Caption". In other cases there is no direct mapping as elements may be missing in one schema, or one element may overlap with two elements in the other schema. This situation requires decisions about best mapping procedure.

Metadata schemas specify how applications should render the metadata. However, applications do not always follow these specifications exactly; this can be due to unclear documentation of the metadata schema, or deliberate implementation decisions. Such “deficiencies” in the software can entail that the content of a metadata schema is not rendered by an application or that it is rendered differently by different applications. For example, while schemas specify the same data format, e.g. Date: YYYY-MM-DD HH:MM:SS, software may display it and require users to enter it differently, e.g. Adobe Bridge: “3/25/2013”, Nikon ViewNX2: “2013/3/25”, Mac Preview: “March 25, 2013”.

Using schemas in unintended ways to make them more suitable to local needs also makes them less easy for common software to render. This can be described as an issue of metadata granularity versus renderability. Image applications integrated with popular operating systems and software packages can typically render some high-level metadata, such as the creator of the image, but they are not capable of rendering more detailed – granular – metadata. This requires more advanced image applications, which are less widely used. Thus there is a trade-off between metadata granularity and renderability that institutions need to address.

A number of issues were raised during the SCREM project consultation and discussions. One was the fear that embedded data would replace collections management. This is far from the object of the exercise, as the project is based on the recognition that the collections management system is always the source of authoritative data. One important element of the new IPTC SCREM schema proposition is that it provides a web link to an on-line updated source of data.

The fact that embedded data is not persistent was also raised. Referring to the previous point, the embedded data is a snapshot of data at a certain point in time. This does not invalidate the data; where used judiciously it provides information which would otherwise be lacking. Embedded data can be productively used for the upload of legacy images and data onto new DAMS, it can be used to send data to other organizations, and it can be used to label images which may be downloaded or copied and otherwise break free from any associated data. A reference to dynamic authoritative data is critical, but data transfer usually involves static data, whether embedded in the images or listed on a spreadsheet. Embedded data has been shown to be a useful tool in a planned workflow.

Conclusions

Through our work we have learned many lessons and reached sharable conclusions. We list them here as enumerated items.

1. Embedded metadata as a useful tool Through investigating the three related issues in our paper we have shown that embedded metadata can be an important tool for managing digitized images and born-digital archives. Embedded data can enhance long-term access and discovery by creating self-describing objects.

2. Awareness of embedded metadata. Many users and curators of digitized or born-digital material are unaware of the flexibility and strength of embedded metadata, especially beyond the technical metadata expressed as Exif. The need for embedded data has been recognized by many who work with images, but needs to extend to other areas of the overall heritage workflow. Those involved in using embedded data can be a) those interested in data harvesting and data distribution b) those who see the value of embedded data at certain points in the workflow and c) those engaged at curatorial level in collections management data. We believe that a useful work-image-data workflow will involve all three sets of people understanding and engaging in the efforts of the others so that the sector can fully realize the potential offered by production and distribution of digital images and associated data. In particular there is a need to overcome perceptions based on past experience. As we have outlined, labelling and mapping has often been inconsistent, but since the advent of XMP, consistency has improved, semantics have been overhauled, and

there is now sufficient information for sound mapping in a number of areas.

3. Supporting standards. The heritage sector needs to promote IPTC Core and Extension standards as the best available standards for embedding data in the image file for public distribution. As software clients, heritage organizations are in a position to influence features available in both collections management and DAMS software, a fact which is often underestimated. Once IPTC Core and Extension are better supported by software, adoption of the standards will be more widespread. The sector will also benefit from an engagement in the SCREM project. The part of the project to propose fields to the IPTC is already underway, and the longer and more complex task of identifying fields for use within the sector would benefit from views from a variety of organizations and departments.

4. Separating data about object and image. In a world where objects are often viewed as an online image, it is important that data about the object and the image of the object is recorded separately. To express data about both the image of a work and the work itself IPTC Core and Extension is found to be effective for data embedding. IPTC Extension includes the category "Object/Artwork" to describe the work. However, the schema does not allow for capturing more detailed information about the nature of the work. The VRA schema is capable of this, but it is not widely implemented. The SCREM schema will be useful in allowing users to embed both data about the image, and about the object.

5. The status of embedded metadata. In line with the increased use of registries and other on-line sources of information, embedded metadata should point to authoritative sources of collections management data. This is increasingly possible through the use of DOIs and RDF as the data in various elements.

6. Working with embedded metadata in practice. It is possible to create data workflows which can operate even when software support is not ideal. The work done by KBDK shows that ways can be found to display important data by concatenating data into one field that is more generally available. A good understanding of data workflows and standards can enable users to flow data into images in ways which display information in ways which can be currently used as well as formatting data to make use of improvements in the future.

7. Education and Outreach. Efforts need to be made to educate both producers and users of metadata about the benefits of using embedded metadata, the standards and tools available, and the ways the data can be used to inform and educate both specialist users and the public. The lack of common tools that interpret granular embedded metadata is a reason to lobby for the use of IPTC Core and Extension and for implementing this photo metadata standard for embedded metadata in more applications.

8. Uptake. Although many people think that embedded metadata needs to be added one image file at a time, plug-ins for Adobe Bridge allow batch export and import of data. This can automate populating files with metadata and also with the perceived limitation of 'fixed embedded metadata'. For large collections of digital images, this allows for metadata updates.

References

[1] Adobe TIFF Specification 6.0: http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf

[2] Exif version 2.3: http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf

[3] IPTC Core and Extension Schema: http://www.iptc.org/site/Photo_Metadata/IPTC_Core_&_Extension/

[4] Dublin Core Metadata Initiative: Dublin http://dublincore.org/

[5] Embedded Metadata Manifesto, http://www.embeddedmetadata.org/

[6] Metadata Working Group: http://www.metadataworkinggroup.com/

[7] Electric Lane, About images and the words that go with them (blog): http://electriclane.blogspot.co.uk/2013/08/heritage-metadata-at-iptc.html

[8] Picture Licensing Universal System (PLUS), http://www.useplus.com/aboutplus/about_coalition_detail.asp?cid=131348256398

[9] MetaShotPpt: http://metadatadeluxe.pbworks.com/w/page/65714041/MetaShotPpt

[10] See: Engineering the Future of the Past blog: http://libraries.mit.edu/sites/digital-archives/

[11] J. Warda, (Ed.) et al, The AIC Guide to Digital Photography and Conservation Documentation, 2nd ed. (2011).

[12] EMDaWG, Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images, Smithsonian Institution, (2010) http://www.digitizationguidelines.gov/guidelines/GuidelinesEmbeddedMetadata.pdf.

[13] Making Pictures available in the Long Now: is embedding an answer?, Greg Reser and Kari Smith, presented at the SAA Research Forum, 2009.

[14] B. Haslhofer and W. Klas, A Survey of Techniques for Achieving Metadata Interoperability, ACM Computing Surveys, Vol. 42, No. 2, Article 7 (2010).

[15] Isaac, Antoine, Robina Clayphan, and Bernhard Haslhofer, Europeana: Moving to Linked Open Data. Information Standards Quarterly, 24(2/3):34-40 (2012), http://www.niso.org/publications/isq/2012/v24no2-3/isaac/

[16] Green, H. E; Saylor, N.; Courtney, A., Beyond the scanned image: assessing scholarly uses of digital collections, Digital Humanities 2013 Conference, http://dh2013.unl.edu/abstracts/ab-230.html

[17] BitCurator Tools. http://www.bitcurator.net

[18] ExifTool (http://www.sno.phy.queensu.ca/~phil/exiftool/); Jeffrey's Exif Viewer (http://regex.info/exif.cgi); Embedded Metadata Explorer (http://embedmydata.com/); EMET (http://sourceforge.net/projects/emet/); and Bulk_Extractor (http://www.forensicswiki.org/wiki/Bulk_extractor).

Author Biography

Kari René Smith is the Digital Archivist for Massachusetts Institute of Technology, Institute Archives and Special Collections working in the areas of digital archives, digital forensics, intellectual property, preservation metadata, and developing and refining workflows for born-digital records and archives. She was Chair of the VRA Data Standards committee and a member of the VRA Embedded Metadata Group that produced the VRA Core XMP info panel. She holds a Master of Science in Information from the University of Michigan, School of Information.

Sarah Saunders runs Electric Lane, providing consultancy and training to heritage clients on image and data management and DAM. She is a member of the IPTC Photometadata Working Group and is working with Greg Reser (UC San Diego and VRA Embedded Metadata Group) on IPTC heritage field proposals and the SCREM project.

Ulla Bøgvad Kejser holds a PhD in Conservation-Restoration Science from the Royal Danish Academy of Fine Arts, School of Conservation. She is a preservation specialist at the Royal Library in Denmark working in the areas of preservation imaging, preservation planning, and curation of both physical and digital collections, especially images.

Acknowledgements
The authors give full appreciation and thanks to Greg Reser without whom this group would not have been formed and this paper not written.