



Archiving and Management of Digital Images Based on an Embedded Metadata Framework

Liu, Cjien-cheng

National Taiwan Normal University, Taiwan
695150038@ntnu.edu.tw

Chen, Chao-chen

National Taiwan Normal University, Taiwan
cc4073@ntnu.edu.tw



國立臺灣師範大學圖書資訊學研究所
Library & Information Studies
NTNU Graduate Institute

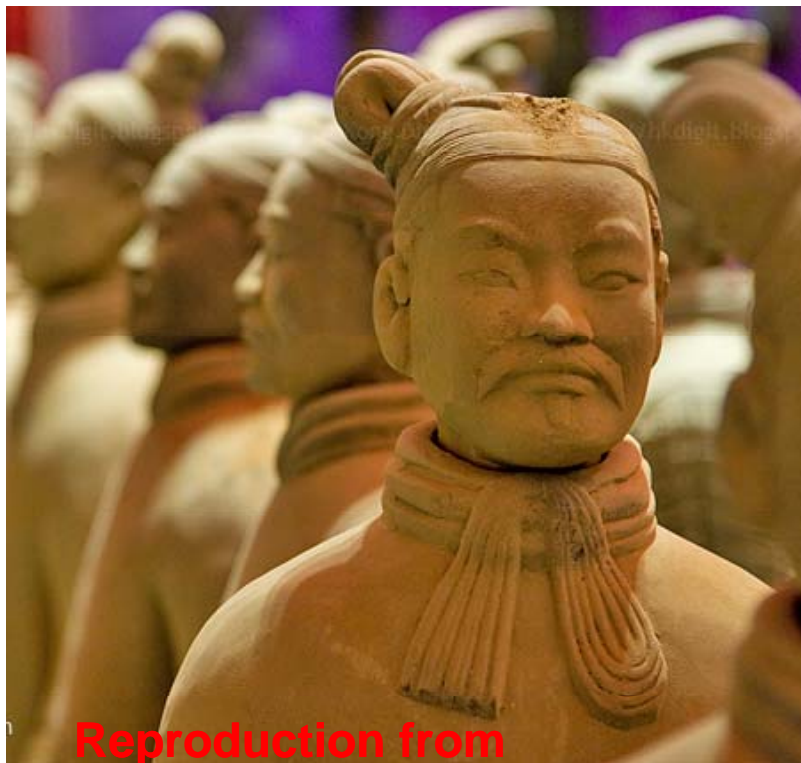


Agenda

- **About Metadata in Digital Preservation**
- **Current image file data management**
- **What is Embedded Metadata?**
- **How does Embedded Metadata work?**
- **Embedded Metadata Standards**
- **Why we prefer XMP**
- **Embedded Metadata Framework and System Design**
- **Demo**
- **Discussion and Conclusion**
- **Q & A**

A little test

- Can you see any difference between these two pictures?



Reproduction from
Terracotta Army exhibit of Hong Kong.

Real terracotta
figure from
Terracotta Army in
Xi'an, China.



About Metadata in Digital Preservation

- **Metadata is very importance information in digital archiving and management.**
- **Metadata enables us to understand the content of a digital image file.**
- **Besides providing content information, metadata also has three other functions:**
 - **Intellectual property & rights**
 - **Access control use**
 - **Search/browse and data exchange use**

About Metadata in Digital Preservation

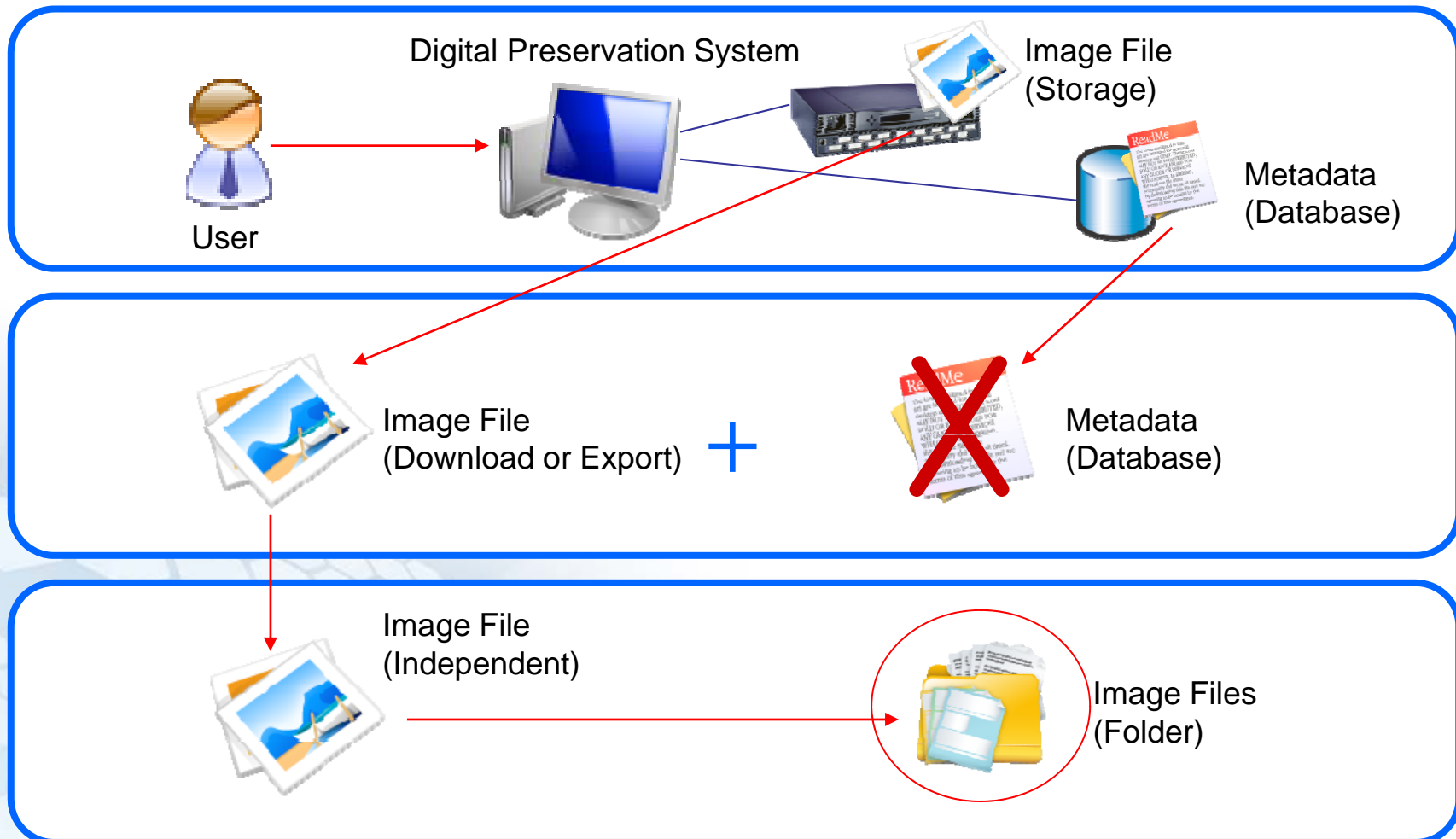
- Metadata & Digital Object (File)



Identifier	Classic_2009080812350_1.jpg
Title	Terracotta Army (兵馬俑)
Creator	Photo by Tommy Wu
Original date of Photo	2009/8/8
Source	Digital Still Camera
Keywords	Terracotta Army; Terracotta Warriors; Terracotta figures;
Abstract	The Terracotta Army (兵馬俑) or Terracotta Warriors and Horses is a collection of 8,099 life-size Chinese terra cotta figures of warriors and horses located near the Mausoleum of the First Qin Emperor (秦始皇陵). The figures were discovered in 1974 near Xi'an, Shaanxi province, China. People believe that the terracotta warriors were based on true people as every face has different facial features and expressions.
Location	Xi'an, Shaanxi province, China.
Rights	National Palace Museum

Current image file data management

- User scenario



Current image file data management

- **User expectations**

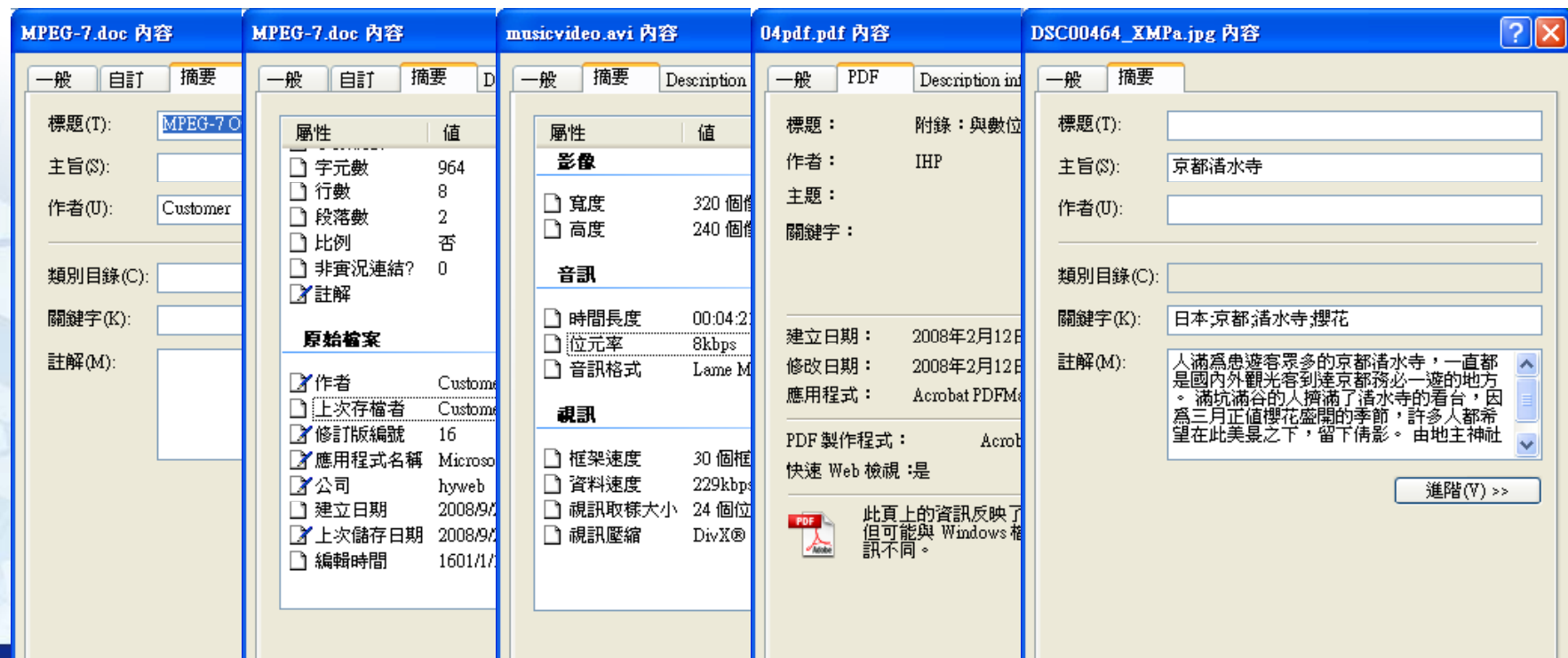
- **User-Friendly and quick reading function.**

- **User's cognitive information structure.**

- **File Name** => **Title** ✓
- **File Type** => **Caption** ✓
- **File Date** => **Creation Date** ✓
- **File Size** => **Max. export size** ✓
- **File Path** => **Category** ✓

Current image file data management

- Adding information in digital files
 - Some file formats can add extra information
 - Tiff, Jpeg, Png, PDF, Word, Excel, WMV, AVI...
 - Windows OS – File abstract preview



What is Embedded Metadata?

- **File Properties**

- **File Name, File Format, File Size...**(including in each file)

- **File Information**

- **Difference Dimensions, Resolutions, Color space...**(according of media types and file format)
- **About media type and file format**

- **Embedded Metadata**

- **User defines and adds information:**
 - **Title, Creator, Abstract, Keywords...**
 - **Caption, Categories, Contact Info, Rights...**

What is Embedded Metadata?

- File Info**

Media Type	File Info	Example	File Format
Images	Resolution (Width, Height), Color Mode, Color Deep, Compress rating, Count of Pages.	800x600, RGB, 8bit, JPEG	TIFF, JPEG, BMP, GIF, PNG...
Documents	Layout size, Pages, Include fonts	A4(210x297mm), 10Pages	Word DOC, PDF...
Multimedia: Film	Dimension (Width, Height), Film length, Frames per Second, Sampling rating, Compress rating, Encode/Decode	720x480, 00:01:25, 30fps, 256kbps, MPEG 2	WMV, MPEG, AVI...
Multimedia: Audio	Audio length, Sampling rating, Compress rating, Encode/Decode	00:03:38, 44.1khz, Stereo, MP3	WMA, MP3, WAV...

How does Embedded Metadata work?

- **Stored Data**
 - **Encode Data (Binary)**
 - **Plain Text (Text or XML)**
- **Stored Section**
 - **Header, Body, Expansion**
- **Access and Stored Method**
 - **Tag**
 - **Flag (Pilot)**
 - **Fixed Field**

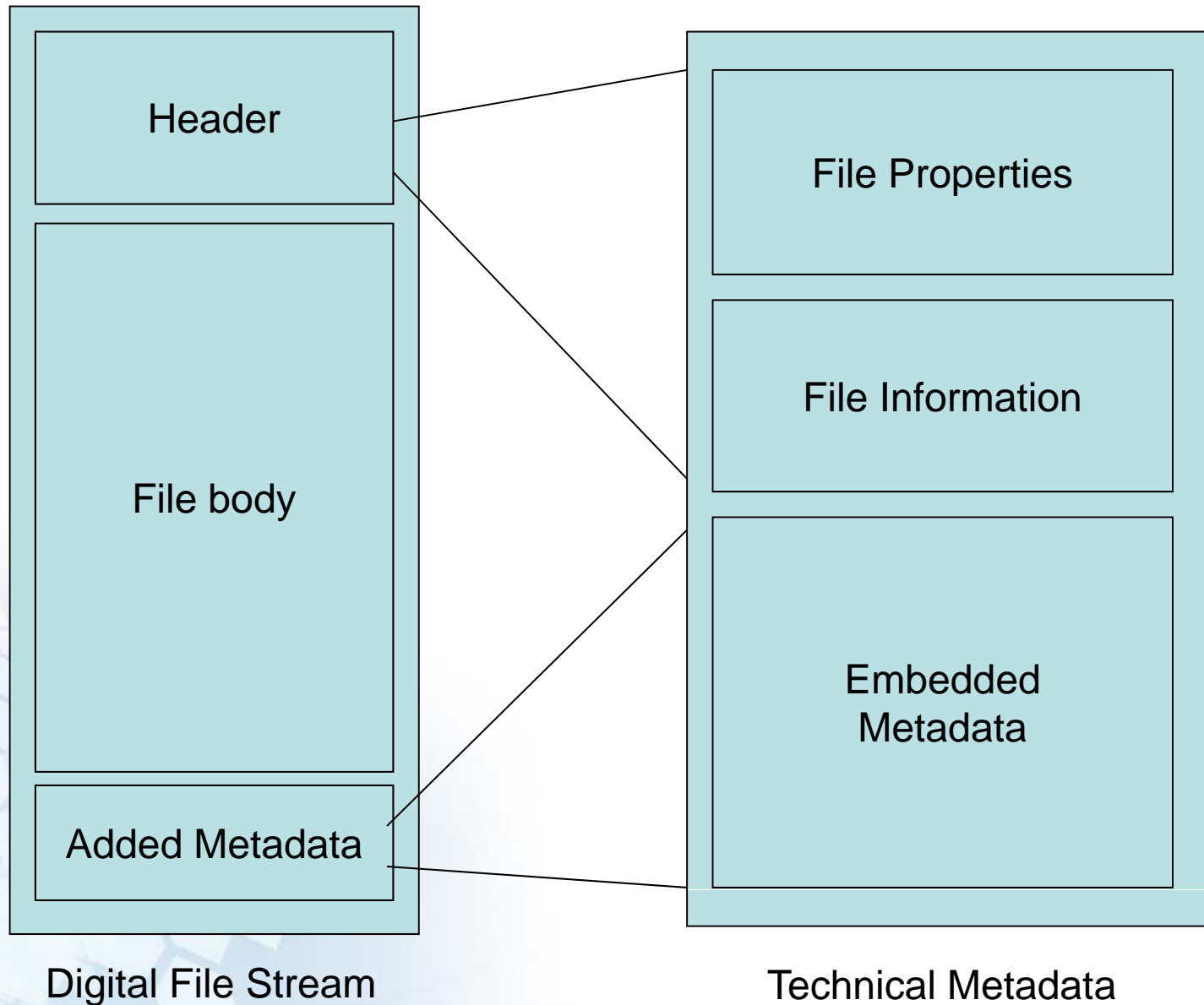
TIFFv6 (ISO12234-2:2001)

Tag ID	Tag Name	Value
256 (100.H)	ImageWidth	800
257 (101.H)	ImageLength	600
258 (102.H)	BitsPerSample	8,8,8
259 (103.H)	Compression	7
274 (112.H)	Orientation	1
282 (11A.H)	XResolution	96
283 (11B.H)	YResolution	96


Code ID (DEC)	Code ID (Hex)	Name	Type	Default	Value
34850	0x8822	ExposureProgram	Short	0	3 = Aperture Priority
40961	0xA001	ColorSpace	Short		1 = sRGB

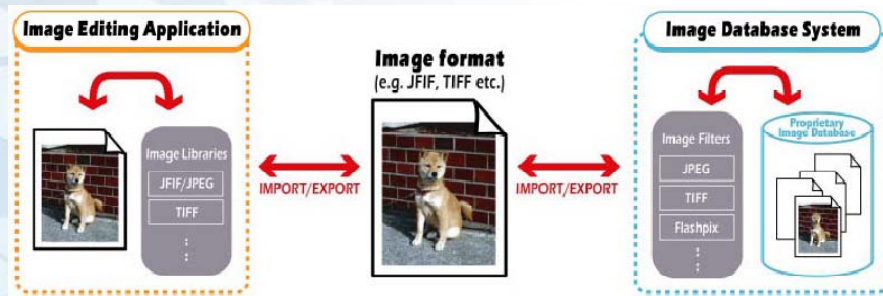
EXIF 2.2 (JEITA CP-3451)

How does Embedded Metadata work?

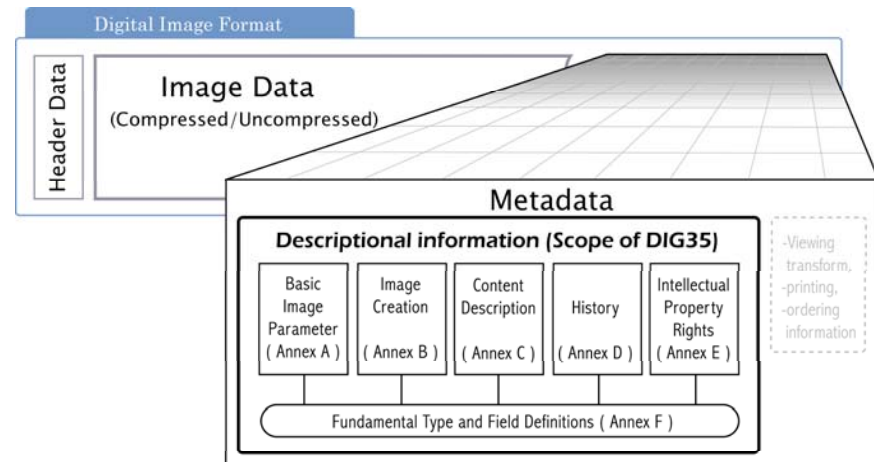


Embedded Metadata Standards

- **DIG35**  (1997)
 - Digital Imaging Group (DIG)
 - 1st Metadata Standard for Digital Images
 - DIG35 Specification v1.0 (2000)



DIG35 Image data interchange model (DIG35, 2000)





DIG35 Digital Image Format. (DIG35, 2000)


Embedded Metadata Standards

- **MIX** *MIX*
 - MIX (Metadata for Images in XML Standard)
 - Library of Congress - Network Development and MARC Standards Office
 - MIX Schema Version 0.2 (2004), 1.0 (2006)
- **ANSI/NISO Z39.87**
 - NISO Technical Metadata for Digital Still Images Standards Committee (2000)
 - Data Dictionary - Technical Metadata for Digital Still Images, ANSI/NISO Z39.87-2006 (2006)


Embedded Metadata Standards

- **DISC**  **(2003)**
 - Digital Image Submission Criteria, DISC
 - International Digital Enterprise Alliance, (IDEAlliance) 
A Working Group of IDEAlliance®
International Digital Enterprise Alliance
 - Domain: Press and Publishing
 - DISC Specification (2007)
 - Digital Images Metadata
 - Contact Info, Special Instructions, Job Identifier, Rights and File History


Embedded Metadata Standards

- **EXIF** The logo for Exif Print, featuring the word 'Exif' in a blue serif font and 'Print' in a blue sans-serif font, with a series of orange dots above the 'i' in 'Print'.
- **Exchange Image Format**
- **Domain: Digital Still Camera**
- **Defines camera-specific metadata**
 - **EXIF 1.0 JEIDA (1996)**
 - **EXIF 2.0, 2.1 (JEIDA-49-1998) (1998)**
 - **EXIF 2.2 (JEITA CP-3451) (2002)**

Embedded Metadata Standards

- **IPTC** 
 - **International Press Telecommunications Council (IPTC) - IPTC Standards Committee**
 - **Domain: News and Press**
 - **IIM (Information Interchange Model) (1991)**
 - **Metadata Taxonomies for the News Industry**
 - **NewsML, NewsCodes**
 - **IPTC Core Schema for XMP v1.0 (2004)**
 - **IPTC Photo Metadata (2008)**

Embedded Metadata Standards

- **XMP**  **(2000)**
Adding Intelligence to Media
 - Adobe eXtensible Metadata Platform (XMP)
 - Domain: Digital Image Process
 - XML/RDF
 - Adobe XMP namespace
 - Adobe Custom File Info Panel (2003)
 - IPTC Core schema for XMP (2004)

Embedded Metadata Standards

- **Adobe XMP characteristics**

- **Stored an XML file in file**

Special
Character

Identify Code

```

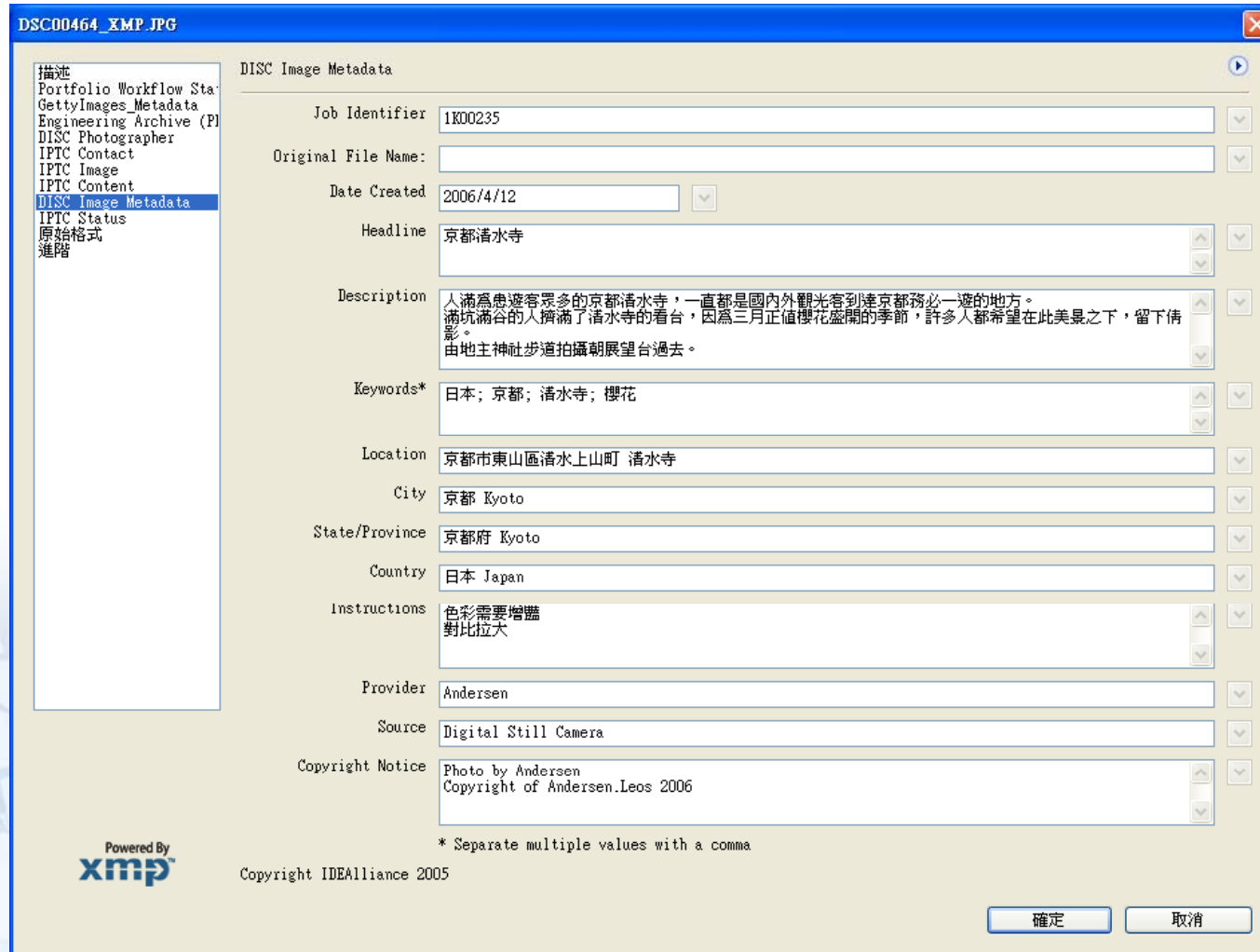
<?xpacket begin="■"
id="W5M0MpCehiHzreSzNTczkc9d"?>
... the serialized XMP as described above: ...
<x:xmpmeta xmlns:x="adobe:ns:meta/">
<rdf:RDF xmlns:rdf= ...>
...
</rdf:RDF>
</x:xmpmeta>
... XML whitespace as padding ...
<?xpacket end="w"?>
  
```

Read and Write
Setting

Why we prefer XMP

- **Specific rules, definition and synchronized data design**
- **Unicode and XML/RDF Standards**
- **Customize and Expand Metadata**
- **User-friendly Operating Panel**
- **Open SDK and Toolkit**

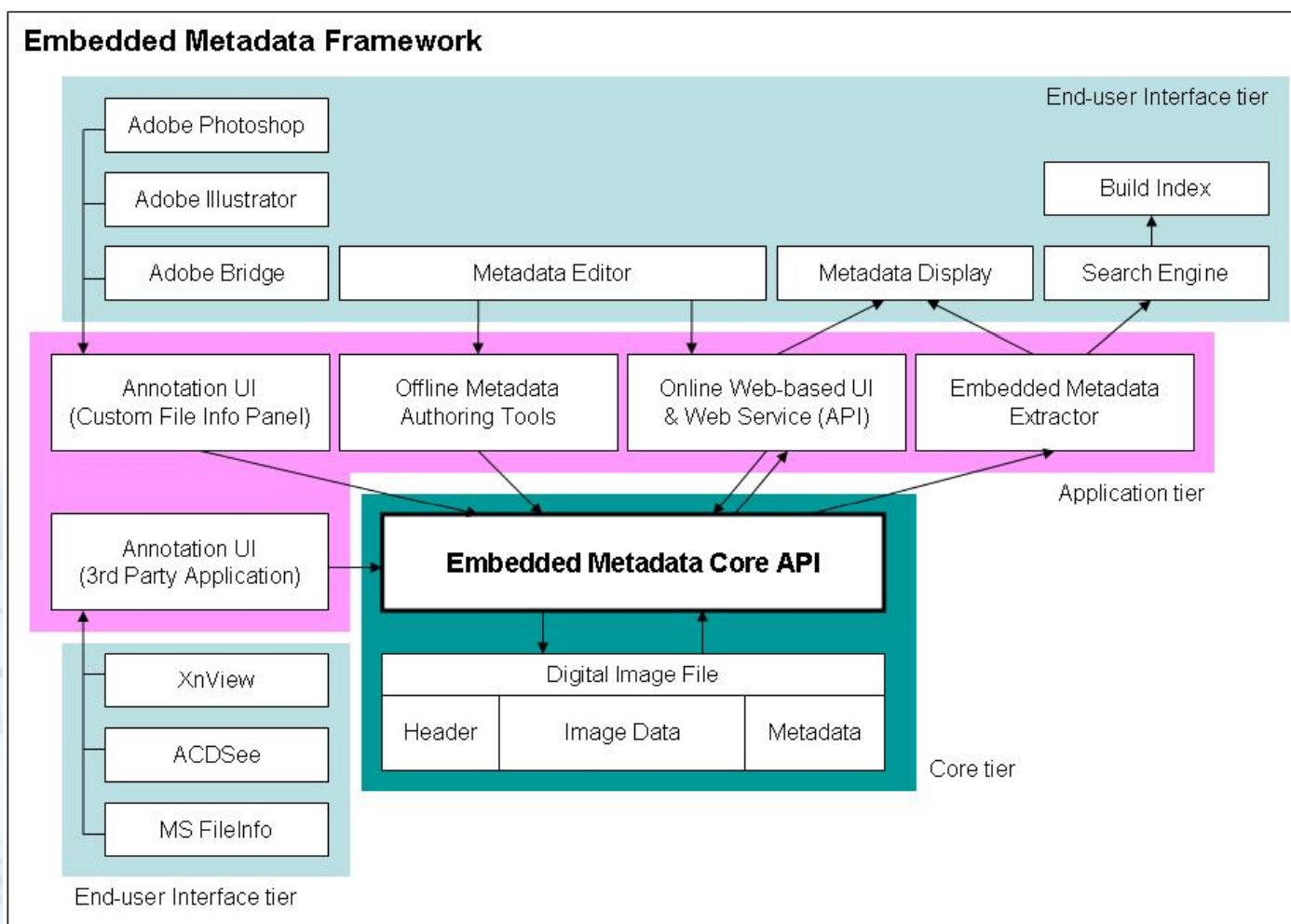
Why we prefer XMP



Adobe XMP Custom File Info Panel

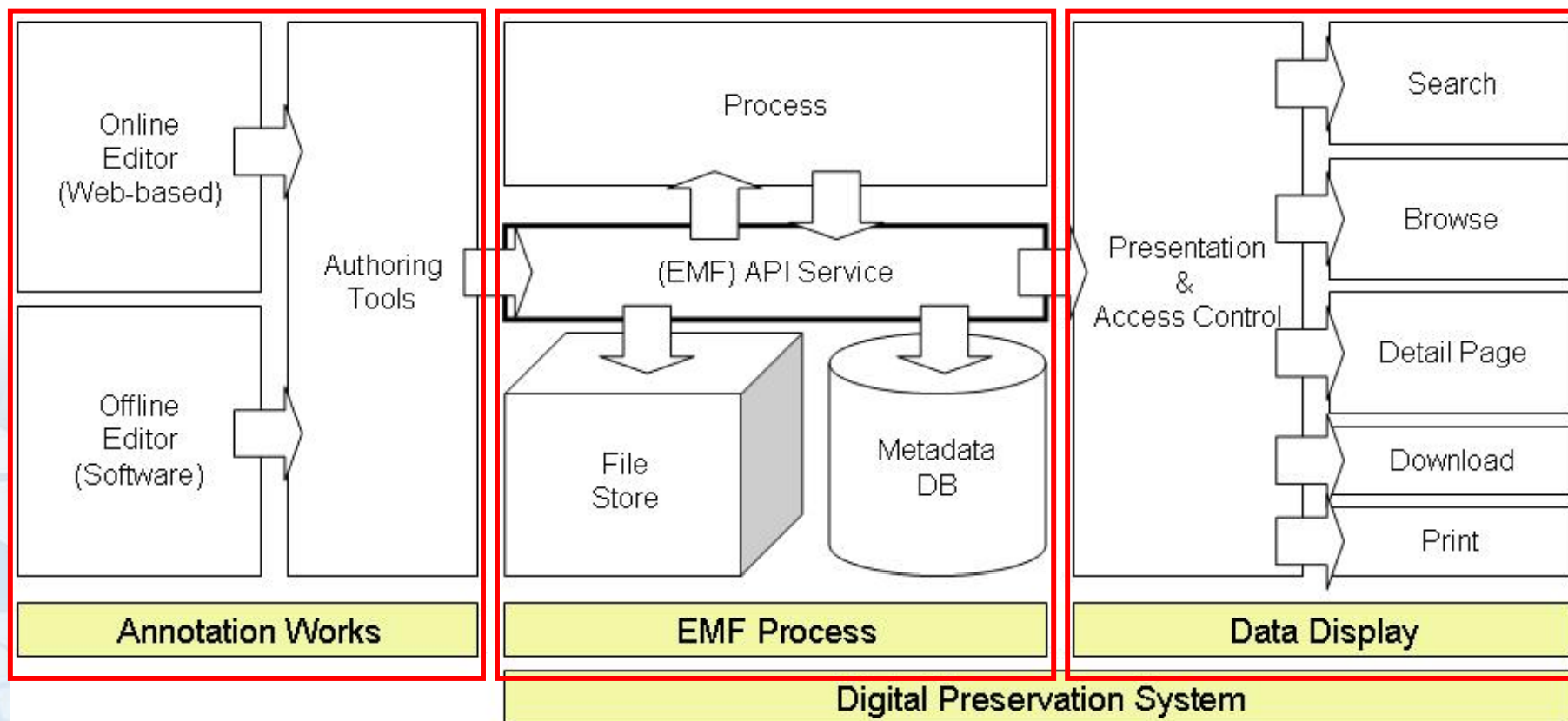
Embedded Metadata Framework and System Design

- Framework concept



Embedded Metadata Framework and System Design

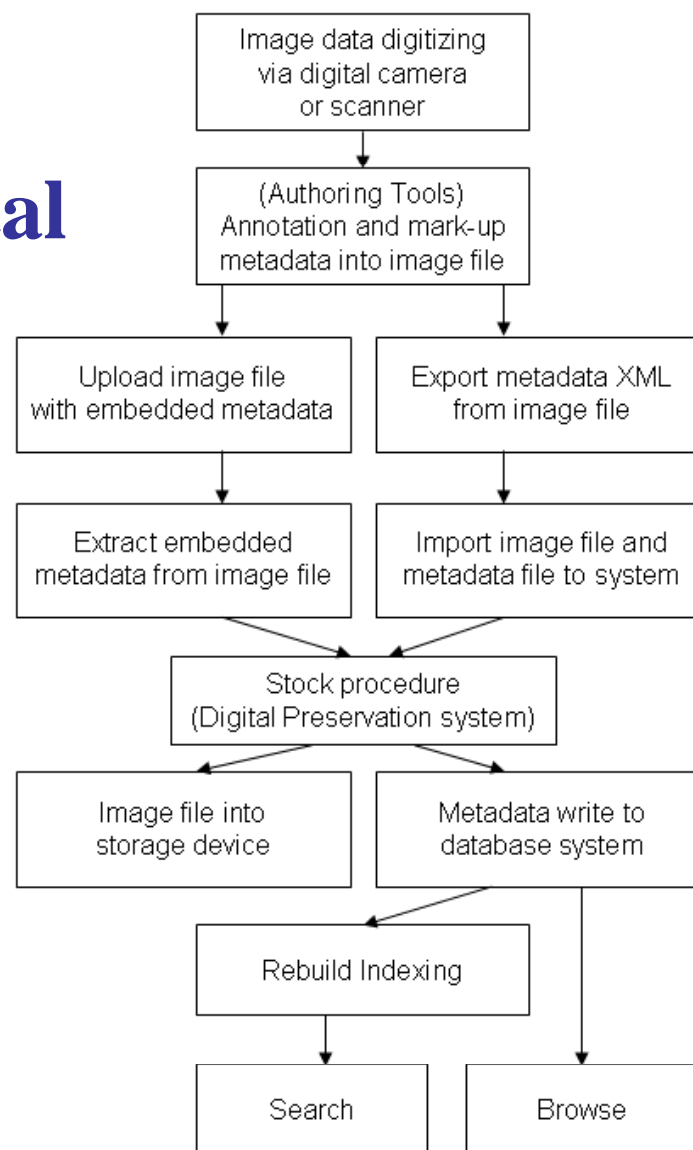
- System Architecture



Embedded Metadata Framework and System Design

- Operating and Working Flow

- Add Metadata into digital image file
- Extract Metadata from digital image file
- Store in DP system
- Metadata => DB
- Image file => Storage
- Browsing / Searching
- Display and Download



Prototype UI

- EMF Metadata display

[Login and Upload](#)
[Browsing and Loading](#)
[Manual and Guide](#)
[Contact Us](#)
[Home](#)

Embedded Metadata Extractor - Image Show

[Image Zone]



[Download](#)
[Zoom In](#)

[File View]

Access File:

[Metadata Zone]

Preview Edit zh_TW Show EXIF

Image Title:	遊客眾多的京都清水寺
Subject Heading:	京都清水寺
Creator:	Andersen.Leos
Creator Role:	Photographer
Caption:	人滿為患遊客眾多的京都清水寺，一直都是國內外觀光客到達京都務必一遊的地方。滿坑滿谷的人擠滿了清水寺的看台，因為三月正值櫻花盛開的季節，許多人都希望在此美景之下，留下倩影。由地主神社步道拍攝朝展望台過去。
Keywords:	日本; 京都; 清水寺; 櫻花;
Category:	int
Sub Categories:	gen; art; trl;
Rights:	Photo by Andersen Copyright of Andersen.Leos 2006
Source:	Digital Still Camera
Digitized Date:	2006/4/12
City:	京都 Kyoto
State/Prov:	京都府 Kyoto
Country:	日本 Japan
Job ID:	1K00235
Instructions:	色彩需要增豔對比拉大
Meta. Writer:	安圖生
Credit:	Andersen

[File Properties]

File Name:	DSC00464_XMPc.jpg
File Format:	JPEG 影像 (image/jpeg)
File Size:	3,713,822 Bytes
Create Date:	2009/9/14 上午 02:58:49
Resolutions:	2592 x 1944 (pixels)
Color Mode:	RGB
Color Depth:	True colors
Multi-Pages:	1

Online/Offline Authoring Tools

Preview	Edit	zh_TW	Show EXIF
Image Title:	遊客眾多的京都清水寺		
Subject Heading:	京都清水寺		
Creator:	Andersen.Leos		
Creator Role:	Photographer		
Caption:	人滿為患遊客眾多的京都清水寺，一直都是國內外觀光客到達京都務必一遊的地方。滿坑滿谷的人擠滿了清水寺的看台，因為三月正值櫻花盛開的季節，許多		
Keywords:	日本；京都；清水寺；櫻花；		
Category:	int		
Sub Categories:	gen; art; trl;		
Rights:	Photo by Andersen Copyright of Andersen.Leos 2006		
Source:	Digital Still Camera		
Digitized Date:	2006/4/12		
City:	京都 Kyoto		
State/Prov:	京都府 Kyoto		
Country:	日本 Japan		
Job ID:	1K00235		
Instructions:	色彩需要增豔對比拉大		
Meta. Writer:	安圖生		
Credit:	Andersen		
		Reset	Save

DSC00464_XMPa.jpg - Info

Description Keywords Origin Advanced Details

Add title, author, description and copyright details to your images.

Title: 遊客眾多的京都清水寺

Author: Andersen.Leos

Description: 人滿為患遊客眾多的京都清水寺，一直都是國內外觀光客到達京都務必一遊的地方。滿坑滿谷的人擠滿了清水寺的看台，因為三月正值櫻花盛開的季節，許多人都希望在此美景之下，留下倩影

Desc. Writer: 安圖生

Copyright: Generate copyright notice automatically

Photo by AndersenCopyright of Andersen.Leos 2006

URL: http://leoslab.net/rights/docid?12345678

1 of 1
(0 items excluded)

JPEG image
2592 x 1944
DSC-F828
2006/4/12 下午 02:49:50
3.53 MB

D:\WebPage\XMPDemo\
UploadFile

Edit as a Collection

OK Cancel Apply Revert

Microsoft FileInfo Tool

Web-based Metadata edit interface



DEMO



Discussion and Conclusion

- **Discussion**
 - **Character-set and encoding**
 - **Data limit and file size**
 - **Multiple and repeatable data**
 - **Metadata security**

Discussion and Conclusion

- **Conclusion**

- **EMF system: Reduces metadata loss**
- **User define: Custom File Info Panel and Metadata schema**
- **Based on XML and Unicode, Easy to exchange**

- **Future**

- **GeoTag**
- **Extend to Multimedia**
 - **ID3v2**
 - **MPEG 7**



GeoTag mark-up and interactive map display (Google Maps, 2009)



Thanks

Q & A

Reference

- **Bibliography 1**

- **Vogl, Howard. (2005). The Use of Technical Metadata in Still Digital Imaging by the Newspaper Industry. Retrieved October 25, 2008, from <https://ritdml.rit.edu/dspace/bitstream/1850/1108/8/HVoglThesis072005.pdf>.**
- **Halaschek-Wiener, Christian, Andrew Schain, Michael Grove, Bijan Parsia, and Jim Hendler. (2005). Management of Digital Images on the Semantic Web.**
- **Halaschek-Wiener, Christian, Jennifer Golbeck, Andrew Schain, Michael Grove, Bijan Parsia, and Jim Hendler. (2005). PhotoStuff – An Image Annotation Tool for the Semantic Web. Retrieved October 27, 2008, from http://www.mindswap.org/~chris/publications/PhotoStuffCR_pid83.pdf.**
- **Hunter, Jane., Zhimin Zhan. (1999). An Indexing and Querying System for Online Images Based on the PNG Format and Embedded Metadata. Retrieved December 2, 2008, from <http://www.itee.uq.edu.au/~ereseach/papers/1999/PNG.pdf>.**
- **OCLC/RLG Working Group. (2001). Preservation Metadata for Digital Objects: A Review of the State of the Art. Retrieved October 25, 2008, from http://www.oclc.org/research/pmwg/presmeta_wp.pdf.**
- **DIG35. (2000). DIG35 Specification – Metadata for Digital Images. Retrieved November 27, 2008, from <http://xml.coverpages.org/FU-Berlin-DIG35-v10-Sept00.pdf>.**

Reference

- **Bibliography 2**

- **ANSI/NISO. (2006). Data Dictionary – Technical Metadata for Digital Still Images. Retrieved November, 11, 2008, from http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-87-2006.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hM9t9qad1BrrORLqssvegis%3D.**
- **Library of Congress. (2008). NISO Metadata for Images in XML Schema. <http://www.loc.gov/standards/mix/>.**
- **IDEAlliance. (2009). DISC, PRISM, AdsML. Retrieved from http://www.idealliance.org/industry_resources.**
- **IPTC. (2005). IPTC Core Schema for XMP. Retrieved October 25, 2008, from http://www.iptc.org/std/Iptc4xmpCore/1.0/specification/Iptc4xmpCore_1.0-spec-XMPSchema_8.pdf.**
- **IPTC. (2005). “IPTC Core” Schema for XMP - Custom Panels User Guide. Retrieved October 25, 2008, from http://www.iptc.org/std/Iptc4xmpCore/1.0/documentation/Iptc4xmpCore_1.0-doc-CpanelsUserGuide_13.pdf.**
- **IPTC (2008). IPTC Information Interchange Model (IIM) - IIM Schema for XMP Specification Version 1.0. Retrieved November 4, 2008, from http://www.iptc.org/std/IIM/4.1/specification/IPTC-IIM-Schema4XMP-1.0-spec_1.pdf.**
- **IPTC. (2008). IPTC Standard Photo Metadata 2008 – IPTC Core Specification Version 1.1 and IPTC Extension Specification Version 1.0. Retrieved November 3, 2008, from http://www.iptc.org/std/photometadata/2008/specification/IPTC-PhotoMetadata-2008_2.pdf.**

Reference

- **Bibliography 3**

- **Adobe. (2005). Extensible Metadata Platform (XMP) Specification – Adding Intelligence to Media. Retrieved August 22, 2008, http://www.adobe.com/devnet/xmp/pdfs/xmp_specification.pdf.**
- **Adobe. (2008). XMP FileInfo SDK – Programmer’s Guide. Retrieved October 8, 2008, from <http://download.macromedia.com/pub/developer/xmp/sdk/XMP-FileInfo-SDK-4.4.2.zip>. (XMP FileInfo SDK.pdf)**
- **Adobe. (1992). TIFF Revision 6.0 (TIFF 6.0 Specification). Retrieved December 29, 2008, from <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>.**
- **DCMI. (2009). Dublin Core Metadata Initiative. Retrieved from <http://dublincore.org/>.**
- **Unicode. (2009). Unicode. Retrieved from <http://www.unicode.org/>.**
- **W3C. (2009). Extensible Markup Language (XML). Retrieved from <http://www.w3.org/XML/>.**
- **W3C. (2009). Resource Description Framework (RDF). Retrieved from <http://www.w3.org/RDF/>.**
- **W3C. (2009). Namespaces in XML 1.1. Retrieved from <http://www.w3.org/TR/xml-names11/>.**
- **IETF. (1996). IETF RFC 2046, Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types. Retrieved October 8, 2008, from <http://www.ietf.org/rfc/rfc2046.txt>.**
- **EXIF.org. (2009). Exchangeable Image File Format (EXIF). Retrieved from <http://www.exif.org/>.**
- **Martínez, José M. (2004). MPEG-7 Overview (version 10). Retrieved December 2, 2008, from <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.**
- **ID3.org. (2009). IDv2 tag for MP3. Retrieved from <http://www.id3.org/>.**
- **SourceForge.net . (2009). GeoTag. Retrieved from <http://geotag.sourceforge.net/>.**