



Resource handout for: "Indexing" Non-Text Assets

This PDF also posted online at: <http://www.controlledvocabulary.com/sla/sla-chi.pdf>

Photometadata.org MetaTutorials and Metadata "Field" Guide:

<http://www.photometadata.org/META-Tutorials> or <http://j.mp/6wUGp>

Web based tutorials, as well as downloadable PDFs and some videos showing you step-by-step how to add metadata using Adobe Photoshop, Adobe Bridge, Expression Media and Photo Mechanic software applications.

<http://www.photometadata.org/META-Resources-Field-Guide-to-Metadata> or <http://j.mp/AuOOW>

An alphabetically ordered list of all IPTC & IPTC Core field label names and their definitions.

Controlled Vocabulary Resources: Metalogging & Filenaming

<http://www.controlledvocabulary.com/metalogging/> or <http://j.mp/ICorY9>

Short one-page tutorials covering how to write effective captions and keywords to aid in the retrieval of images from a Digital Asset Management (DAM) system.

<http://www.controlledvocabulary.com/imagetatabases/filenaming.html> or <http://j.mp/95dvTs>

Recommendations and limitations for filenames applied to digital assets.

<http://www.controlledvocabulary.com/imagetatabases/cv-photo-mechanic-code-replacement.html>

Using Code Replacement & Variables in Photo Mechanic to add keywords from a spreadsheet.

Jeffrey's Online Metadata Viewer:

<http://regex.info/exif.cgi>

View embedded metadata for a wide variety of image files without leaving your browser.

Metadata Manifesto:

<http://metadatamanifesto.blogspot.com/> or <http://j.mp/mdOfns>

A blog in support of adopting guiding principles, standards and technology to promote image metadata use.

Embedded Metadata Manifesto on the IPTC website:

<http://www.embeddedmetadata.org/>

An updated version of above, which encourages the use of embedded metadata in a wider variety of digital files.

<http://www.iptc.org/std-dev/photometadata/examples/microdataPages/demoWebpage-Microdata.html>

Example of how embedded metadata can become "microdata" which can be indexed by search engines.

CEPIC Workflow docs:

<http://www.iptc.org/goto?imagemetadatahandbook>

A set of documents covering how to use and assign metadata in an image driven workflow.

ExifTool (and ExifTool GUI):

<http://www.sno.phy.queensu.ca/~phil/exiftool/>

Command line tool & Windows GUI for reading & writing IPTC, Exif, & XMP metadata to many digital files.

Image Keywording Services

<https://imagecategorizer.services.open.xerox.com/>

<http://open.xerox.com/Services/imaging-demos>

Xerox system trained to auto-categorize 706 subjects; and which allows you to test a selection of your images.

<http://imense.com/>

Commercial keywording service Imense has couple of tools, like Annotator, or Autotagger worth testing.

©2012 David Riecks/ControlledVocabulary.com and used with permission. For more information on Controlled Vocabularies and how to embed keywords as Metadata using any of a dozen different imaging applications or databases please visit <http://www.ControlledVocabulary.com/products/>

What is a Controlled Vocabulary, and how is it useful?

Takes the Guess Work out of Searching

A controlled vocabulary makes a database easier to search. Since we have many different ways of describing concepts, drawing all of these terms together under a single word or phrase in a database makes searching the database more efficient as it eliminates guess work. However, arriving at this efficiency requires *consistency* on the part of the individual indexing the database and the use of *pre-determined* terms.

A Familiar Concept

It's likely you are already familiar with the concept of controlled vocabulary. Phonebook Yellow Page listings are arranged using controlled vocabulary. For example, a search for "Car Dealers" leads you to a note to "*see Automobile Dealers.*" At a basic level, this is how a controlled vocabulary system works.

One Search is All it Takes

Conducting a search in a database that uses controlled vocabulary or indexing terms is efficient and precise. The biggest advantage to controlled vocabulary is that once you do find the correct term, most of the information you need is grouped together in one place, saving you the time of having to search under all of the other synonyms for that term.

Finding a Balance

It's difficult to say whether controlled vocabulary or natural language systems give the best retrieval performance. Free Text or Natural Language systems often provide more results in a shorter time span because you are searching all the fields of a given database (the Google search engine is a form of free text search). Such searches work well for very specific searches, however, when a topic is older or broader in scope, you likely will retrieve irrelevant hits. You also may miss some records relevant to your search because you didn't choose the proper search term. As with a web search, searching a database requires striking a balance between preciseness and generating enough hits to make the search successful.

Stop Words

In many online databases you should keep in mind that there are certain words that are ignored. These are called "Stop Words." Common stop words are words such as **the, a, an, this, and that**. While stop words may provide some useful content in Natural Language Processing; most keyword based approaches do not use grammar to parse user input, so this content is not used effectively.

Much of the information on the ControlledVocabulary website concerns how to apply a controlled vocabulary used to describe images within an [imagedatabase](#). Within the "[metallogging](#)" section you will find resources and suggestions on how to efficiently [caption](#) and [keyword](#) your images. If you would like to join others in discussing this topic further, just enter your email address in the sign-up box on the forum sign-up page (<http://www.controlledvocabulary.com/forum.html>).

David Riecks founded ControlledVocabulary.com as a resource to help others learn how best to build controlled vocabulary lists, thesauri, and keyword hierarchies for describing images in databases. He serves as the Chief Technical Advisor for the PLUS Coalition, which developed the Picture Licensing Universal Standards. In 2004, Riecks helped to revise the decade old IPTC standard for use with Adobe's new XMP format. He continues to serve on the IPTC Photo Metadata working group which released the updates to the IPTC Core and IPTC Extension schemas. He previously chaired the SAA Imaging Technology Standards and was the Photo Metadata Education project leader (which received an award from the US Library of Congress to create Photometadata.org). In recognition of his years of work to further photo metadata standards, he was named as a "Pioneer of Digital Preservation" by the Library of Congress in 2009.

©2012 David Riecks/ControlledVocabulary.com and used with permission. For more information on Controlled Vocabularies and how to embed keywords as Metadata using any of a dozen different imaging applications or databases please visit <http://www.ControlledVocabulary.com/products/>