

“Putting Words to the Image” — Keywording:

Best Practices to Accommodate Distributors

A report from the CEPIC Seminar in Malta, June 2008
by David Riecks

This CEPIC session included key personnel from all of the major stock image distributors: Getty Images, Corbis, Jupiter Images and Alamy. The focus of the presentations and discussion was on streamlining how contributors can best prepare their images for submission, especially in regards to the submission of keywords. As nearly all of the speakers notes, good keywords are essential for delivering a good search experience for customers and in some cases—such as the Getty Images (creative and PunchStock) websites—they are the only word terms that are searchable. On the Alamy, and Corbis websites visitors are allowed to search the caption/description field of the image, but these fields are not given as much weight as the keywords in the search algorithm. By contrast, the Jupiter Images site not only allows the site visitor to search the caption but gives this field more weight than the keywords.

Mary Forster, the Director of Search Data Strategy for Getty Images, stated that the key is to “put the right content in front of the customer, so they can have accurate and relevant search results.” Forster explained that they give all content / keyword service providers a 46 page document called the “Getty Images Guidelines for Keyword Providers of Creative Content” that explains all the details. They expect the content provider to specify the coverage of basic areas, location, concepts, people and ages of models, and even objects, except in the case of studio shots where there is now relevant cultural content.

When stating ages, care should be taken to be accurate. If the model is 22 years of age, then use the “20-24” bracket, and don’t use any additional brackets beyond the one that encompasses the actual age of the model.

Keyword providers, however, should refrain from including tangential, speculative or opposite keywords. If you have a picture of a stethoscope, do not include other keywords like doctor, or heart disease. Forster said that they do use their own hierarchical controlled vocabulary as it adds a lot of value to their search capabilities, as synonyms are tied to the preferred terms.

They don’t want partners to include synonyms and variants. That’s what their system does. Any synonyms they add serve as backups or alternatives to the main terms, and also include translations into other languages. Forster said that Getty Images expects 100 percent compliance; and content will be rejected if it doesn’t meet these specified requirements. They do, however, offer training to new providers if requested.

Andrew LaBonte, the Senior Manager of Search Metadata at Corbis said that they have their own “Distributor Keywording Guidelines” document for contributors.



CEPIC

Centre of the Picture Industry

CEPIC-OFFICE

Fritschestraße 22
10585 Berlin
Germany
Tel. +49.30.88 91 01-60
Fax +49.30.88 91 01-61
ceplic@ceplic.org
www.ceplic.org

However, they currently do most of their own keywording in-house using a hybrid, “Mapalogging” system. This highly efficient system allows each keyworder to catalog approximately 1000 images per week, rather than the 400 they were achieving previously. They do review partner images and remove any inappropriate keywords. For instance, any form of “Keyword spamming” is not tolerated, such as the example he showed with an image showing only the top half of a young freckle-faced girl with a frog on her head against an orange background. She was not wearing any clothing that could be seen, yet the image had the term “casual wear” as well as “summer” added to it as keywords.



At present, on average, approximately 39 percent of their media partners submitted keywords end up being rejected.

It’s important, said LaBonte, to note that the types of keywords needed will vary according to the type of stock photography. For editorial content, captions are essential and the date is very important. For travel imagery, the date is less important, and the location information (city, state, country) is of paramount importance.

CEPIC
Centre of the Picture Industry

Alan Capel, the Head of content for Alamy, said that the number of images grows every day to the point that even he has to check the site to see how many images they have. At the time of the conference that number was approximately 12 million pictures.

CEPIC-OFFICE
Fritschestraße 22
10585 Berlin
Germany
Tel. +49.30.88 91 01-60
Fax +49.30.88 91 01-61
ceplic@ceplic.org
www.ceplic.org

Unlike the other distributors, Alamy does require that contributors do all their own keywording and caption. However, they do not supply a Controlled Vocabulary. Instead they require contributors to actually group the keywords they add into three different keywords fields— essential, main and comprehensive—in order to help weight which words are most important. As the sort order of images is determined by customer activity, the metadata provided by contributors plays an important part.

There is substantial competition as everyone is doing their best to have their images seen. The actual sort order of images on site is driven by customer activity. If your image appears in the search results and is ignored, you lose. If it appears and is clicked on, you win, and if it is licensed it you really win.

The Alamy search formula gives a very high weighting to “Essential” keywords, followed by the “Main” keywords, then the “Comprehensive” and finally the Description field. Currently they only extract the IPTC “Keyword” and “Description / Caption” fields, so contributors typically need to manually enter, or break up the keywords into these three keyword fields on their site. It’s not necessary to separate each keyword with a comma; you only need to add a space between each word. Of course this also means that their search engine is not particularly useful for searching for multi-word keywords or phrases.

The click-through-rates are measured for all pictures and are also used to help determine which keywords are most relevant. In order to keep things running smoothly they won’t ask you to update an image, if you exceed the specific character lengths, instead your image will be rejected, and you will have to resubmit later. Capel recommended that contributors use the ‘LEN’ function in excel to check field lengths to avoid this from happening.

Janice Lodato, the Director of Search Data for Jupiter Images, noted that they currently prepare images for one of eight or nine search engines. She described how they keyword, starting with the “Aboutness” of an image, and then expanding to include concepts, age and ethnicity, appearance and action, and location

information. They ask that their partners use a simple relevance test and ask themselves, if they were searching for "X" term would they want this image in the results. In their future she sees a need for normalization, standardization and more automation.

Unlike the other image distributors in the panel discussion, Jupiter Images is one of the few that requires that contributors submit their metadata in a tab-delimited/text data file (Excel preferred), with Keywords in a comma-delimited data file (Excel preferred) and delivered as part of metadata file. In addition, if the contributor also has the keywords translated, those should also be included as a comma-delimited data file (Excel preferred).

Kirsti OSullivan, owner of Keywording.com, gave some great tips on how to use special spreadsheet functions to ease the pain of preparing submissions. Functions such as concatenate, allow you to join together columns of data, whereas the LEN option allows you to know the exact character length of a field. In addition, the use of Find/Replace options can ease the pain of correcting a consistent misspelling within a spreadsheet of image submission entries.

In the Question and Answer segment, the audience posed questions regarding differences in how distributors deal with American and UK spelling, conceptual keywords, building controlled vocabularies, and the use of IPTC metadata fields for input into systems.

David Riecks
June 2008

--

David Riecks (that's "i" before "e", but the "e" is silent)
david@riecks.com <http://www.riecks.com/>
Midwest/Chicago ASMP



CEPIC

Centre of the Picture Industry

CEPIC-OFFICE

Fritschestraße 22
10585 Berlin
Germany
Tel. +49.30.88 91 01-60
Fax +49.30.88 91 01-61
ceplic@ceplic.org
www.ceplic.org