



Extensis Portfolio™

VERSION 3.0

**DEVELOPING A
KEYWORDING STRATEGY**

Developing a Keywording Strategy

This section describes how to organize your catalogs with keywords to speed image retrieval. By learning how to develop an effective keywording strategy you will:

- Make your multimedia catalogs easier to work with and more efficient to use.
- Decrease the possibility of “losing” catalog items because of inadequate keywording.

About terminology: Both indexing and keywording refer to the process of guiding a user to the contents and location of print or electronic information. These terms are used interchangeably throughout this section.

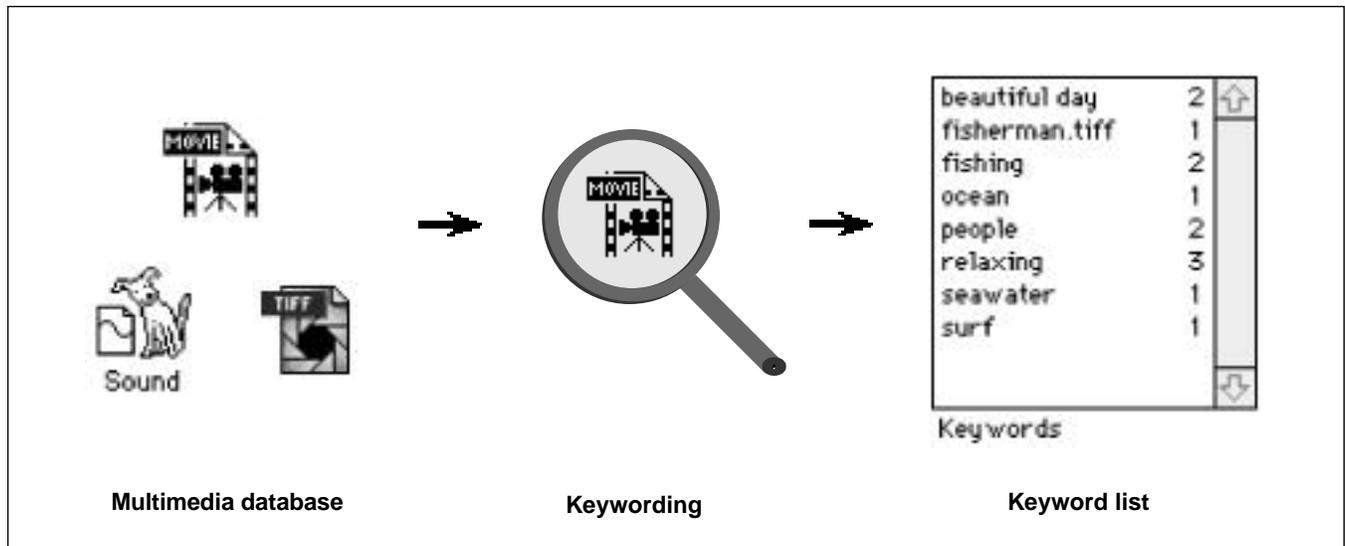
The Importance of Keywords

It is possible to use Portfolio without keywording: simply allow it to process your collection and use the miniature images to guide you to the items that you need. However, the main benefit of having a catalog is the ability to find things as quickly and easily as possible. As your collection grows, and especially if you are in a shared work situation, you will want to use keywording strategies with Portfolio to enjoy all the benefits of an efficient catalog.

It takes time to build an indexing scheme and to actually perform the task of indexing. But if you do not take the time or keyword haphazardly, you will create an inadequate index, which is akin to having no index at all.

The direct relationship between keywording and searching makes it critical to develop meaningful keywords that make retrieving items faster and easier. In this chapter, you will be introduced to two approaches to keywording: natural language and controlled vocabulary. Later sections describe how to use keywords in Portfolio.

Keywording a multimedia database



Choosing a keywording approach

Before you begin to enter keywords into your catalog, you need to define an approach that works for you.

Library scientists have developed two approaches to keywording—natural language and controlled vocabulary. In natural language keywording, terms are freely assigned to items. In controlled vocabulary, keywording terms are clustered, and preferred terms are assigned to items.

Both approaches offer their own advantages and disadvantages. The approach that you choose depends on practical concerns, such as time, collection size, and personnel. These two approaches to keywording are compared on the following page.

Natural language keywording—advantages & disadvantages

Type of user	Advantages	Disadvantages
Large organization	<p>Cost is distributed.</p> <p>New words are added as they emerge.</p> <p>Searching is fast and easy.</p>	<p>Too many people involved in keywording resulting in too many keywords per item.</p> <p>Keyword list can get too big, chaotic, and eventually useless.</p> <p>User must think of the keywords others have assigned in order to retrieve relevant catalog items. User must know the vocabulary of the subject discipline when searching.</p>
Small organization	<p>New keywords may grow slowly since catalog may grow slower.</p> <p>New words are added as they emerge.</p> <p>Only requires 1 or 2 people to be involved in the process.</p> <p>Searching is fast and easy.</p>	<p>May get lax about documenting decisions.</p> <p>May not gather a list of keywords.</p> <p>May not use keyword list regularly (it's like starting over again every time).</p>

Controlled vocabulary keywording—advantages & disadvantages

Type of user	Advantages	Disadvantages
Large organizations with many users and/or a large catalog OR small organizations with a large collection.	<p>Reduces the uncontrolled use of words or phrases.</p>	<p>Not as specific as natural language keywording.</p>
	<p>Ensures consistency and uniformity in the use of keywords.</p> <p>Shows hierarchical and parallel relationships among terms through cross-referencing.</p> <p>Solves many semantic problems and combines terms into a single keyword.</p> <p>Cost is distributed.</p>	<p>Some users may feel that it hampers creativity.</p> <p>Some users may feel constrained by use of approved keywords when other terms may be more natural.</p> <p>Requires more time, effort, and (possibly) personnel to construct.</p>

Using natural language keywording

Natural language keywording is the easier approach since you assign keywords freely. You simply look at the item, then assign keywords which “naturally” occur to you.

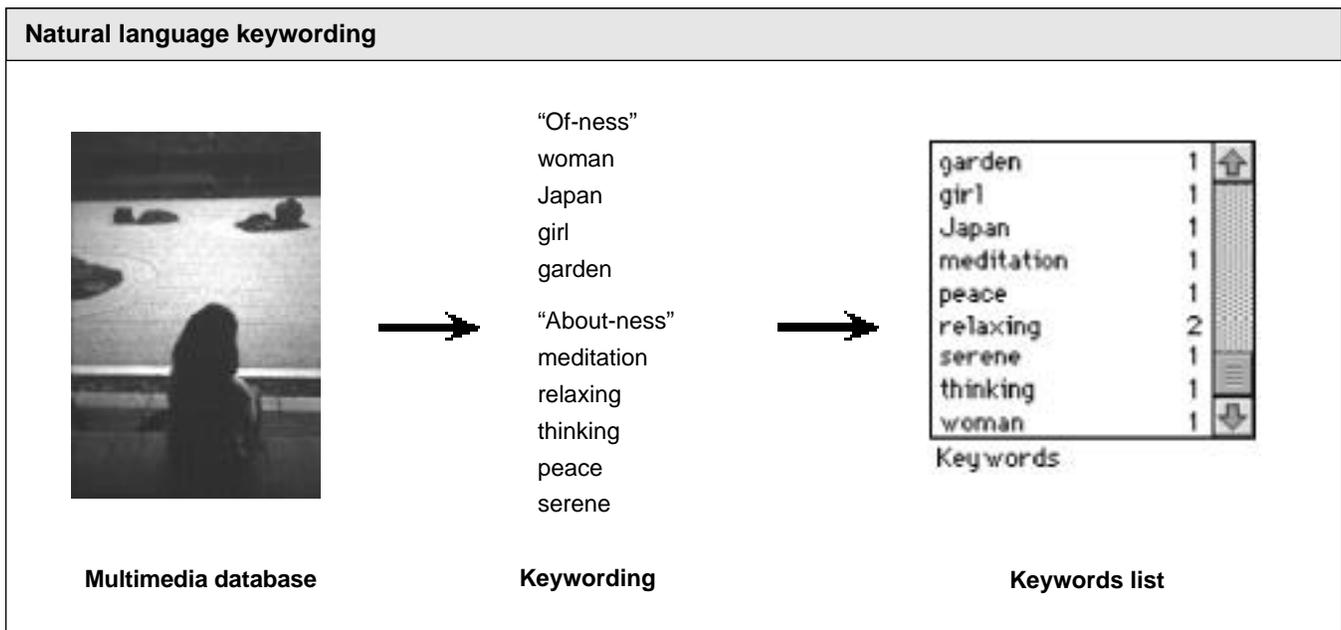
The process of performing subject analysis on multimedia is still a new and developing area. Current research indicates that subject analysis of visual works or images can be done by asking yourself two broad questions: “What is this an image *of*?” and “What is this image *about*?” Your answers to these questions become your keywords.

Defining “of-ness” and “about-ness”

“What is this an image of?” or its “of-ness” deals with the *factual* meaning of the image. “Of-ness” describes people, places, objects, conditions, and actions. This makes “of-ness” easy to keyword—it’s a woman, a bird, the sky.

“What is this image about?” or its “about-ness” is the *expressional* meaning of the image. “About-ness” deals with mood, emotions, and abstract concepts. Is the image about happiness? Is the image about peace? These subjective concerns make “about-ness” more difficult to keyword.

For a complete explanation on how to keyword images, refer to Sara Shatford’s article, “Analyzing the subject of a picture: a theoretical approach,” listed in the bibliography at the end of this chapter.



Using controlled vocabularies

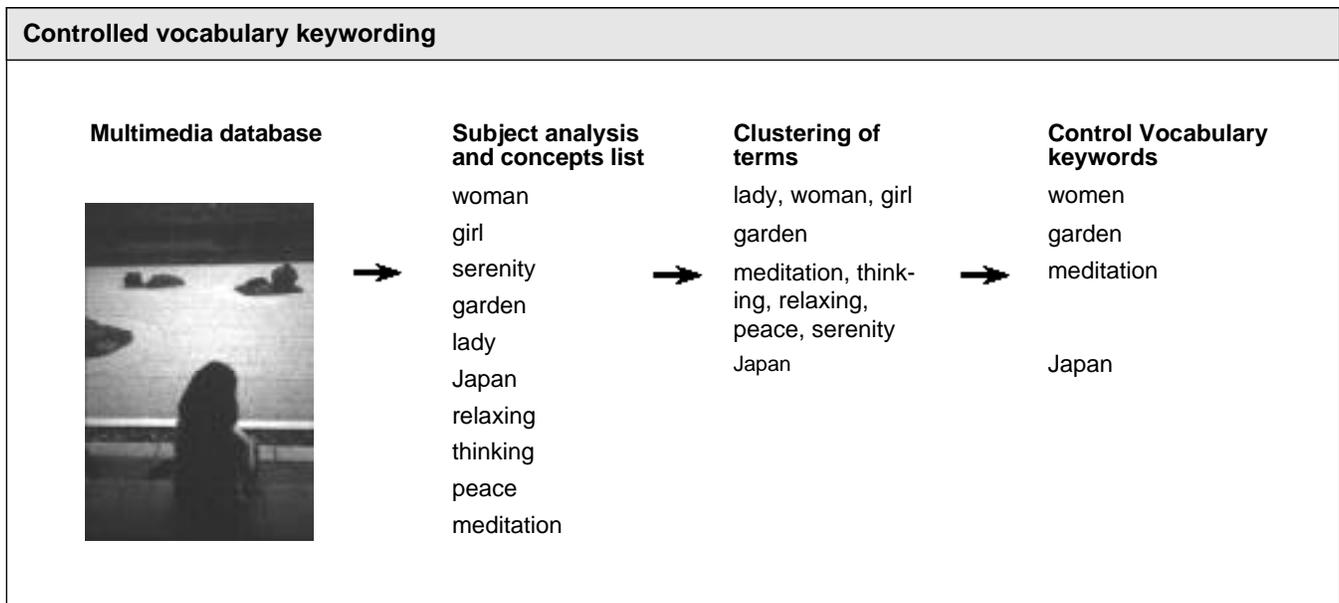
A controlled vocabulary takes natural language keywording one step further: controlled vocabularies

- Designate what keywords can be used by creating an approved list for keywording or searching.
- Help you locate terms that best describe an object when keywording or searching.
- Guide you if you cannot think of a broader or narrower term to use when keywording or searching.

Constructing a controlled vocabulary

The process of constructing a controlled vocabulary is performed outside of Portfolio and consists of three steps. The first step is subject analysis of your catalog. Look at each of the images and identify its “of-ness” and “about-ness” as if you were doing natural language keywording.

The second step is to establish a concepts list. This list consists of the terms you defined during subject analysis. If you were doing natural language keywording, these terms would become your keywords in Portfolio. With a controlled vocabulary, they are an intermediate step to defining the actual Portfolio keywords.



The final step involves clustering the terms on your concepts list into groups of word associations and then choosing a term to define each group (girl, lady, woman = women). This defining term (women) becomes the accepted keyword and the terms within the group become your controlled vocabulary.

Before clustering your concepts list, keep in mind “user warrant.” User warrant means that a term is warranted (valid) if users will use it to locate information. Get to know what terms your users might employ to locate information. Use terms that are familiar to them and common to the subject area.

While you are clustering your terms, there are several questions you’ll want to ask before deciding on a final keyword:

- Is there an exact term available?
- Is there a narrower term that might be used?
- Should a broader term be used?
- Should a related term be used?
- Should two or more terms in the keyword list be coordinated?

Assign as many keywords as it takes to completely capture all aspects of the subject. You might want to establish a policy about the number of keywords that can be assigned to any given item.

Approaches to clustering terms into a controlled vocabulary

Method	Most applicable to...
Create a committee of people who know who the users are and can suggest keywords based on this knowledge and experience.*	Large organizations with the time, money, and personnel to devote to this approach.
Review the current literature and reference sources of the subject area for keyword ideas.	Smaller organizations whose catalogs grow at a moderate rate, and those that can have a non-subject specialist perform the literature review and select the keywords.
Use an existing keyword list in the appropriate subject area.	Organizations, such as those in the sciences or business professions, may already have keywording languages. For example, graphic designers might use the <i>LC Thesaurus for Graphic Materials</i> published by the Library of Congress as a source for their keywords. You can find such indexes at either public or academic libraries.

* If the items in the catalog are highly specialized (such as a catalog of medical images), committee members should consist of people who not only know who the users are, but are also expert in the subject matter.

Structuring your controlled vocabulary

Clustering terms results in a word list or thesaurus of keywords that reflects the subject matter of the database. The keywords in a thesaurus have a prescribed format (which is slightly different from the Roget-type of thesaurus you might be familiar with).

A key feature of a thesaurus is the hierarchical or term relationships among words. Adding broader, narrower, and related terms as well as cross-references to your controlled vocabulary will help direct users to the accepted keywords.

Establishing term relationships

In the example, Women represents the keyword. Keywords are the user's entry point into the word list.

A Scope note (SN) reference is used to restrict or to clarify the use of a term. Although not quite a dictionary definition, the scope note should be a brief description of what you mean by that term.

B Broader term (BT) reference refers to a term hierarchically broader in concept than the keyword.

C Narrower term (NT) reference refers to a term hierarchically narrower in concept than the keyword.

D Related term (RT) refers to terms that show no hierarchical relationship to the keyword. Related terms may have their own entry and can be cross-referenced on the word list.

Not all of these term relationships need be present for each term. The term relationships can be added at any time as they become necessary.

Women

SN (Scope Note): here are entered catalog items about the female gender

BT (Broader term): female

NT (Narrower term): girl

RT (Related term): lady

The following additional references will help direct users to preferred terms:

Used for (UF) refers to synonyms or variant forms of the preferred term. Used for means that the term **cannot** be used as the keyword. UF is also used to lead the user to more general terms.

For example:

DOGS
(keyword that can be used)

UF GERMAN SHEPHERDS
(cannot be used)

The reciprocal of **used for** is **use** and it is used as a cross-reference to a preferred term from one that is not allowed. For example:

Cross-stitch USE Needlework

Defining syntax for your controlled vocabulary

As the controlled word list is generated, certain syntax rules should be observed. Consult *American National Standard Institute Guidelines for Thesaurus Structure, Construction, and Use* for a complete description of syntax rules.

Syntax rule for headings

Rule	Examples
Use the plural for concrete concepts.	"Bytes" NOT "Byte," "Pixels" NOT "Pixel"
Use the singular for abstract concepts.	"Religion," "War"
Use gerunds for activities.	"Relaxing," "Eating"
Use qualifiers to differentiate between homonyms.	"Locks (hardware)" versus "Locks (hydraulic engineering)"
Use a phrase if a concept cannot be adequately expressed by a single term.	"Cultural background," "Cell biology"
State compound terms in their natural order, not inverted.	"Cultural background" NOT "Background, cultural"

Displaying and arranging a keyword list

Regardless of your keywording approach, your final keyword list should be accessible to every user. If you are using a controlled vocabulary, the keyword list must be available in printed form to establish and show term relationships.

Your keyword list should be user friendly. Individual units, keywords, and references should be quickly recognizable. You can export the keyword list and then format it using a word-processing application. Use bolding, indents, and different fonts to differentiate between keywords and term relationships.

Partial keyword list from a controlled vocabulary word list

Keyword list page 25

Women

SN: *Here are entered catalog items about the female gender*

BT: Female

NT: Girl

RT: Lady

Using keywords with Portfolio

Portfolio provides a visual language which displays in miniature pictures the contents of a catalog, one screen-full at a time. This thumbnail display is further enhanced with an index. The index is expressed through the master keyword list, found in the Catalog Info window. Much larger volumes can be searched by using keywords, rather than by looking at screen after screen of thumbnail images. The ease and efficiency of a search is increased by using effective keywording strategies.

Keywording procedures

Keywording is simply attaching terms to items to act as identifying markers. The actual choice of words, patterns, and associations for the items depends on the purpose of your collection. Consistency and descriptiveness are good general guidelines to follow when developing keywording strategies for your catalogs.

To begin keywording items in Portfolio:

- 1. Designate a person to be responsible for assigning keywords and general catalog administration.**

- 2. Prepare the materials.**

The source files should be descriptively titled, in appropriately named folders, and recognizable by Portfolio. (File types that Portfolio knows are listed in the Catalog Properties dialog, Options tab.) You can add almost any file type to the list currently recognized by Portfolio (see Adding a file type).

- 3. Add items to the catalog.**

Create a new catalog by launching Portfolio, opening the File menu, and choosing “New;” or open an existing catalog by choosing “Open.” Add new items to a catalog by choosing “Add Items” from the Catalog menu.

- 4. Add basic descriptions.**

Select each thumbnail in the Gallery view and open the “Item Properties” dialog box.

Identify equipment, sources, unabbreviated titles, personnel, and any other important information that you want to keep track of for each individual item.

- 5. Add keywords to groups of items.**

By adding keywords to groups rather than one at a time you can ensure consistent application and spelling of keywords throughout the entire catalog.

Go through the whole collection and gather one large basic group together. For example, to keyword a set of images with the keyword *people*, follow these steps. Create a new Gallery (in the Window menu). Search through the catalog for the appropriate images and

simply drag the desired images into this new gallery window. Once you have all the items you need, select everything (Select All in the Edit menu) and open the Edit Keywords dialog (from the Item menu). Type *people* into the dialog, click the ‘Add’ button, and then click ‘OK’. The keyword has now been added to this set of items

Using similar steps, you can gather from the *people* group smaller groups, such as *children, families, men, women, faces*. The *faces* group can then be further subdivided into different expressions and age groups. Besides *people*, other basic groups might include pictures containing *water, sports, and cars*.

6. Export the catalog’s keyword list.

Select Catalog Properties from the Catalog menu and choose the Statistics tab. Click the ‘Calculate’ button to generate the list of keywords in the catalog. Click the ‘Save Keywords’ button to export the list to a text file. Click ‘OK’ to close the dialog.

7. Add the keywords to the Master Keyword List.

Select the Administration dialog from the Catalog menu. The dialog will open to the Master Keyword List tab. Click the ‘Import’ button and find the file you just saved in the previous step. The keywords will be imported into the Master Keyword List (MKL). Click ‘OK’ to save the changes to the MKL.

8. Open the Master Keyword palette.

Choose Keyword Palette from the View menu. A palette will be displayed showing all the master keywords. To do a search for one of these keywords, simply double-click on the word in the palette. To add this keyword to another item, simply drag the keyword from the palette to the item.

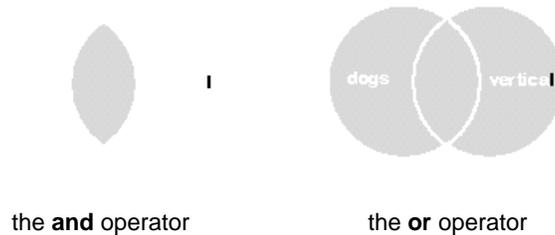
The palette is an excellent tool for the casual user, since it presents a list of all the keywords you want them to see. Instead of having to guess how the catalog has been keyworded, they can simply browse the keyword list in the palette to choose the keyword appropriate to the search.

Tip:

If keywords are added inconsistently, items are lost in the collection. In a search for a picture of a mountain, the desired items might have the keyword *mountain*, others might use the keywords *Mt. Baker, Cascades, summit, peak*; or the picture may be the right one but not have any such keyword at all. You may find it helpful to develop a checklist for processing each item. By adding the same word systematically to all of the relevant items it is possible to locate the desired parts of the collection quickly and efficiently.

Keywording and searching

You will be gathering large groups of related items (like vertical images) as well as specific individual items with similar types of attributes (like pictures of dogs). Boolean searching (using *and* and *or* as search operators) works best when looking for items with broad overlapping attributes.



In the illustrations above, one circle represents all of the items which contain the keyword *dogs* and the other circle represents all of the items which contain the keyword *vertical*. Not all of the vertical images are *dogs*, and not all of the *dogs* are vertical images. The *and* search retrieves only the items that have both terms. The *or* search will retrieve all of the items that contain either keyword, a much larger group.

The screenshot shows a window titled "Find - D32_Test.fdb". It features a "Saved Finds:" dropdown menu set to "Default". Below this, there are three rows of search criteria, each with a dropdown menu for the operator and a text input field for the keyword. The first row has "Keywords" selected, "starts with" as the operator, and "mount" as the keyword. The second row has "and" selected, "Filename" as the operator, and "HI" as the keyword. The third row has "and" selected, "Extension Win" as the operator, and "gif" as the keyword. At the bottom, there are two checkboxes: "Find in gallery" and "Display results in new gallery", both of which are unchecked. Below the checkboxes are four buttons: "More Choices", "Fewer Choices", "Find All", and "Find".

In Portfolio you can use the expanded Find window to define a Boolean search. Type in the first keyword and select “more choices.” Type in the next keyword (most searches use two terms, however, you can combine up to four). Use “does” or “does not” and “start with” “contain” etc., then select either the *and* operator or the *or* operator button to widen or narrow your search. Click “Find” to start the search.

Subject analysis

You can identify an item by its physical features and by its conceptual contents. We recommend that you develop a checklist for processing each item using the following criteria.

Basic item identification

File names: Spelling accuracy is critical to retrieval, especially when titles contain abbreviated words.

Folder names: Portfolio can use the name of a folder containing several items as a keyword for identifying those items. This applies only to the folder immediately containing the items, not to subfolders or larger folders.

Titles: Project titles that can be used to group together many individual components are excellent keywords. Sometimes unabbreviated item titles can work nicely as keywords.

Creator(s): We recommend including the file creator's name as a keyword to provide a link to a body of work. By creator we mean artists, photographers, and authors.

Software: Some programs, such as *Adobe Gallery Effects*, can create both TIFF and PICT file types. It can be useful to distinguish items by the software that was used to create them.

Dates: You must decide on the significance of dates and how they will be expressed if you use them as keywords. For example, all dates could indicate when the item was created, or last modified, or when it was added to the catalog (this date is found in the Info window, Command + I). Any of the above applications of dates makes it possible to look at the collection from a chronological perspective.

Time-based art: The duration of a sound or movie is the best way to distinguish between derivative files or files with similar characteristics. You can find the digital length, format codes and other dimensions in the information window of an item, but you have to re-enter any of these details to use them as keywords. Time can be expressed with a simple prefix, like *t* which will cause all of the keywords that are time lengths to be gathered together in the master keyword list. For example: *t01:15*

Sound files: Depending on how you use sound or music, you may want to identify items by project, duration, effect, function, voice content, source, instrumentation, and by adjectives that describe melody, style, and tempo.

Image orientation: In the stock photography industry, orientation is an important aspect of an image because art directors have very specific needs, such as a *vertical* picture for a magazine cover, or a *square* or *horizontal* image to fit in a particular layout.

Special photograph types: Such as *abstract*, *aerial*, *landscape*, *medical*, *microscopic*, *panoramic*, *portrait*, *telescopic*, *underwater*, etc.

Tip:

Be consistent in how you apply chronological references. You could express copyright dates one way (c1993), creator and actual creation date in another way (DW11/17/93), and dates that items were published in yet another way (NYTimes1/12/93). In these latter examples, the alphabetical display will cause coded keywords to sort together by the c for copyright, by the author's coded initials, and by the publication symbol. By changing the order to Year/Month/Day, you can gather material together by year, or if you place the month first, the material can be gathered by month.

About-ness and of-ness identification

As you develop your subject analysis checklist, several of these generic questions may be applicable to your keywording.

- **Are there people shown?**
Number of people, familial status, age group(s), activity of people, emotions evident.
- **What are the picture's most significant objects or qualities?**
Animals, location, geographic (St Louis), general settings (balcony), season, time of day, mood, signs (such as billboards, words on t-shirts, etc.), dominant color(s).
- **Can situational terms be applied?**
Professional jargon, project related keywords, client related keywords, date related tags for items, specially coded information.

Maintaining the catalog

For consistency and continuity, you may want to designate a catalog administrator who has an understanding of how Portfolio works and a familiarity with the requirements of your particular database.

The catalog administrator

In a situation where there is one primary user, that person will be responsible for the whole catalog and will carry on the keywording and catalog maintenance functions as they deem appropriate.

In a group situation, designate one person to be the catalog administrator or liaison. The catalog liaison will be responsible for developing and maintaining consistent procedures for adding items, descriptions, and keywords and for training users who are searching for items.

Designing the keyword list is the main task. The catalog liaison will have to anticipate the terms that will be prominent, arrange the keywords meaningfully, and decide how to gather all of the related items that belong together.

Start with some natural language keywording and make a few experimental catalogs before settling on an approach. This offers an opportunity to make observations about what is needed from the overall system. As time goes on and the catalog grows, you or your catalog administrator will probably find it advantageous to actively construct the keyword vocabulary.

Continuing research and development of keywording

Programs like Portfolio allow you to develop sophisticated keywording strategies for many very specific uses. As you continue to apply keywords to your catalog, you may want to learn more about keywording theories and methods. For more information consult the bibliography below.

Bibliography

- American National Standards Institute. *American National Standard Guidelines for Thesaurus Structure, Construction, and Use*. New York: American National Standards Institute, 1980.
- Cleveland, Donald B. and Cleveland, Ana D.. *Introduction to Indexing and Abstracting*. Englewood, Colorado: Libraries Unlimited, Inc., 1990.
- Getty Art History Information Program. *Art & Architecture Thesaurus*. New York: Oxford University Press, 1990.
- Library of Congress. *LC Thesaurus for Graphic Materials*. Washington, D.C.: Library of Congress, 1987.
- Shatford, Sara. "Analyzing the Subject of a Picture: A Theoretical Approach." *Cataloging & Classification Quarterly*. 6 (1986) : 39-62.
- Zinkham, H. and Parker, Elisabeth Betz. *Descriptive Terms for Graphic Materials: Genre and Physical Characteristic Headings*. Washington, D.C.: Library of Congress, 1986.