

HRAF Use for global Indexing

著者(英)	Keiko Fukukawa
journal or publication title	Senri Ethnological Reports
volume	28
page range	153-157
year	2002-03-15
URL	http://doi.org/10.15021/00002051

5.7 HRAF Use for global Indexing

*Keiko Fukukawa
National Museum of Ethnology*

Recently, using the Outline of Cultural Materials (OCM), I classified the Mongolian photographs from Konagaya and also, at the request of Yasugi, I tried to retrieve data of Mexican sample materials (artifacts) from the British Museum(BM) and the National Museum of Ethnology (NME). In this article, briefly explaining the Human Relations Area Files (HRAF), I will describe my considerations about classifying the photographs using the OCM and the problems I have found with the differences of the OCM classifications between the BM and the NME for the same types of Mexican sample materials. Therefore, here the OCM classification will be the main theme of discussion.

5.7.1 HRAF Files

The HRAF Files is a collection of file materials on the cultures of the world. The files were developed by the Human Relations Area Files, Inc., an international organization in USA, for the comparative study of the world cultures. The sources of the files are mainly selected ethnographies and the all pages are coded by numbers in accordance with the OCM, HRAF's subject classification system. Moreover, they are filed according to the Outline of the World Cultures (OWC), HRAF's area or cultural group classification. Those two unique systems were designed by the HRAF for the purpose of retrieving world cultural information rapidly and effectively.

5.7.2 OCM (Outline of Cultural Materials)

The OCM deals with the aspects of human behavior (social life, customs, material products, ecological setting, etc.) comprehensively, and divides them into 716 categories. The 716 categories are grouped into 79 major subject divisions, each numbered from 10 to 88. Within each major subject division, a third digit from one up to nine is added and designated by a three-digit number code. For example, the [22: Food Quest] is subdivided into eight categories as follows: {221: Annual Cycle}, {222: Collecting}, {223: Fowling}, {224: Hunting and Trapping}, {225: Marine Hunting}, {226: Fishing}, {227: Fishing Gear}, and {228: Marine Industries}. Each

category has a brief explanation indicating the range of information which may be classified under that category. Obviously, each category has an extensive amount of information that is nearly impossible to briefly explain, so that the explanations are relatively simple outlines. Under each three-digit category, cross-reference categories are listed.

In the Japanese version of the OCM, "Bunka Komoku Bunrui", each original English description under each three-digit category was simply assigned a new five-digit number, consisting of the original three-digit number followed by a decimal and two more digits, from .01 to .99. (e.g., {221.01: Seasonal distribution and succession of food-getting and other economic activities}, {221.02: Seasonal migrations in the gathering or production of food}, ...{221.99: etc.})

5.7.3 OCM Classification of Photographs

My job was to code for the over 180 photographs taken in Mongolia using the five-digit of the Japanese version of the OCM mentioned above. During my working, the first trouble I encountered was not being able to obtain any informational contents immediately from the photographs. To remedy this problem, the photographer soon supplied me with a caption for each photo. According to the captions, I also found that there were three parts in all among the pictures, each with a main title.

With the coding of photograph, non-visual information must usually be consulted with written documents and the coding must be assigned also to information contained within that documents. If there are no documents, it is a general rule to code only for information appeared visually.

Moreover, the coder must resist to mark for their own imagination or implication for the visual information. For example, supposing that there is a photo of an adult woman milking a cow. In that situation, the main subject the coder only can take may be "adult woman", "clothing" "cows" and "milking", not more than this. In other words, the case like assuming "milking cows is a female job" and assigning the category {462: Division of labor by sex} must be avoided. To make inferences and draw conclusions from photographs is not the responsibility of the coder, but that of the researcher.

The second point of consideration in coding is the number of the OCM codes for a single picture. Basically, human behavior has many related aspects, and so, two or more codes can be assigned. In the case of the multiple five-digit codes within a three-digit division, however, the effectiveness of retrieval must be considered. It seems reasonable to assume that you are retrieving HRAF Files using the OCM three-digit codes with the multiple adjacent codes, there is often repetition of information, which

becomes a bothersome waste of time for you. Considering the OCM as a system that classifies various aspects of human behavior into categories, coding should only be done for what will be main subjects, as much as possible. This is because, related information other than that can be coded by the cross-references categories presented under that OCM three-digit categories.

5.7.4 OCM Classification of Sample Materials: Examples from the BM and the NME

In order to try to retrieve using the OCM five-digit codes, I chose four Mexican sample materials, "pottery figurine", "mask", "burner", and "pendant", for which the BM classified according to the OCM, and using those OCM codes, I tried to retrieve the same type of the samples owned by the NME. I would not go into detail here, in any way, the results of my search were zero —pottery figurines, two —masks, zero —incense burners, and zero —pendants. Then I searched those zero —items by their actual names and I found all the same sample materials in the NME just under the different OCM codes. If there has been a simple misapplication of the OCM, it is out of my control, but those fundamental differences of the OCM coding for the same materials has serious meaning.

First, there is a clear difference in data contents between the BM and the NME. For the data from the BM, descriptive reference materials are well equipped. As the result, the OCM coding is relatively detailed and accurate. Each artifact has OCM codes along with information on how it was made, what it was made of/from and even on details on its ornamental design. On the other hand, the NME's OCM coding is dependant exclusively on an artifact's name, the brief explanation on the manner in which it is used, and so on. Compared to the BM's case, given information is overwhelmingly insufficient.

Such a big difference in classification between those two museums could have been avoided by drawing up a set of classification rules. Considering that there are no rules for corders between them at the moment, here I could not further describe the examples.

Incidentally, HRAF has a set of classification rules for analysts of written materials. Within the rules, I will cite some OCM categories below that must be helpful when you classify material cultures.

- The products processed from raw materials
- ex. 28: Leather, Textiles, and Fabrics.
- 32: Processing of Basic Materials.
- 38: Chemical Industries.

39: Capital Foods Industries.

—The raw materials involved and their acquisition

ex. 31: Exploitative Activities.

—The tools and machines employed

ex 40: Machines.

41: Tools and Appliances.

—The description and use of the final products

ex 29: Clothing.

30: Adornment.

34: Structures.

40: Machines.

41: Tools and Appliances.

—Additional considerations must be needed for the social function of an item.

ex Ornaments as status symbols.

301: Ornament.

554: Status, Role, and Prestige.

ex Clothing differences between social classes.

291: Normal Garb.

565: Classes.

(Human Relations Area Files 1965:78-79)

5.7.5 Necessity for Classification Rules (Guide for Corders)

Even though there is an elaborate classification system in place, it is necessary to have rules for coding. In order to use a large quantity of classified and organized materials in a wide range of research, classification rules and the classification procedure must possess a high degree of uniformity and regularity. And the procedure and its rules must be maintained from beginning to end. Even in the case where a rule is biased, it must be consistently biased throughout. If, for example, we do this and then find something that needs to be fixed, we can easily take out all the data with this bias and correct them. Any change to this basic structure will bring about extreme disorder, so we must be very cautious in classifying. Even in the case that a coder is changed, the way of classification will change easily based on the new corder's background.

If you are aiming to integrate the information, you must code, as a general rule, totally different material using the same classification system and rules. However, depending on the nature of the materials, slight changes and improvement of the classification system and rules also becomes necessary. In that case, even with a minor change, we must carefully consider what effects it may have, and we must be careful again to

maintain uniformity and regularity.

5.7.6 Is OCM Five-Digit Classification Effective?

When I classified Mongolian photographs using the OCM five-digit categories, I often found photographs that seemed to fit certain five-digit categories but they actually did not. I unwillingly had to code {99: etc.} for them. To classify under {99} means that there are a lot of information under this ambiguous category and the meaning of "classification" is therefore lost. In other words, effectiveness is lost, and further classification within {99} is needed. In another case regarding five-digit classification for one picture or one artifact, there are multiple five-digit codes under a single three-digit code. Retrieving data from Mexican materials, I recognized five-digit codes as retrieval were just useless for the most part. This means that the OCM is originally a three-digit classification system and five-digit categories were simply divided the three-digit categorical explanation into many smaller parts. In order to be fit for a classification system, the current five-digit categories need to be developed, however, that will take time and energy equal to creating another OCM. Therefore, I may say here that the three-digit OCM classification is more effective than the five-digit for the present.

Finally, to increase the effectiveness of classifying and retrieving data by using the OCM's three-digit classification system, I would like to briefly introduce how a proposed classification is done at another place. In that system, when they classify with a three-digit code, the specific term of a related material's main subject or name is added. For example, in Mongolian case of "cutting ears of sheep", the photo would be coded {233: ear-cutting}. And if one retrieve {233}(a three-digit code), the information classified there is indicated by a combination of the three-digit code and the specific terms. From there, the user can only pick up the information he/she wants (e.g.. ear-cutting) from the specific terms listed under {233}.

In case that such specific terms are allowed to accumulate in the future, it would be easily to find how they could become a thesaurus.

References

Human Relations Area Files
1965 *Analyst's Handbook*.

Kreiss, Laura & Elizabeth Stockton
1980 Using the Outline of Cultural Materials as a Basis for Indexing the Content of Ethnographic Films. *Behavior Science Research* 15:281-293.

