

Alison Gilchrest. Factors Affecting Controlled Vocabulary Usage in Art Museum Information Systems. A Master's paper for the M.S. in I.S. degree. July, 2001. 43 pages. Advisor: Helen R. Tibbo

This paper describes the results of a study designed to assess the state of controlled vocabulary adoption behavior in United States fine art museums. Personnel responsible for collections data at thirty art museums were asked to report on aspects of vocabulary control, staff resources and data utilization within their institutions.

Sixty percent of the museums studied reported using at least one controlled vocabulary reference while entering data, and nearly ninety percent use a customized list of local authority terms. The factors affecting adoption of controlled terminology were found to be: institutional resistance to change, lack of staff time, training, financial and administrative support. In addition, the complexity of the vocabularies themselves and/or their inability to adequately represent the scope of many collections were also cited as barriers to use.

Headings:

Information systems -- Special subjects -- Museums

Information systems -- Special subjects -- Art

Authority control

Museums -- Collection management -- Data processing

FACTORS AFFECTING CONTROLLED VOCABULARY USAGE IN  
ART MUSEUM INFORMATION SYSTEMS

by  
Alison Gilchrest

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Information Science.

Chapel Hill, North Carolina  
July, 2001

Approved by:

---

Advisor

## Table of Contents

Introduction .....	2
Literature Review .....	5
I. Controlled Vocabularies .....	5
II. Controlled Vocabularies in the Museum Context .....	8
III. Controlled Vocabularies in Use: Pilot Projects .....	11
IV. Controlled Vocabularies and the Networked Future .....	15
V. Controlled Vocabulary Usage – Previous Inquiries .....	17
Methodology .....	19
Results .....	21
I. Vocabulary Usage .....	21
II. Museum Resources .....	26
III. Attitudes and Behaviors Toward Collections Management .....	27
IV. Networked Initiatives .....	33
Conclusion .....	34
Appendix A: Interview Protocol .....	37
Appendix B: List of Participating Institutions .....	39
Appendix C: Demographic Breakdown of Sample .....	40
Bibliography .....	41

*In the future the . . . computerized catalog will contain only selected and carefully controlled types and quantities of data such as . . . key word indexes to those categories of data considered significant finding devices.*

Robert G. Chenhall, *Museum Cataloging in the Computer Age*, 1975

## **Introduction**

The last decade has witnessed a significant change in the way museums manage their collection information. Making the leap from card-based or homegrown systems to “off the shelf” collection management databases is now a familiar task from the largest and most prominent collections all the way down to small local historical societies. Consequently, the shift toward robust digital collections management presents an opportunity for institutions to reconsider the depth, accuracy and potential uses of their data, and thus implement systems that will best help them achieve optimal standardization for improved retrieval.

Unlike in-house or custom systems, vendors provide comprehensive data structures, pre-defined functionalities for common museum tasks, end-user training and support. A typical collections management database allows for integration of multiple facets of an object – such as loan status, insurance, conservation history, donor information, descriptive text, artists’ biographical information – all in one place, all accessible in a matter of a few clicks. It also provides sophisticated reporting capabilities, allowing museum staff to generate frequently needed lists and forms in a fraction of the time it would take to prepare them manually.

Not surprisingly, within the past several years vendors have also begun to integrate vocabulary capabilities for indexing and retrieval of collection records, a trend that corresponds to the development of improved vocabularies for cultural heritage information and increased consumer demand for access. Standardized terminology applied to key data fields is a demonstrated way to ensure improved retrieval performance in most database environments, whether it means selecting terms from an established vocabulary tool or agreeing on a set of authority terms to be used and enforced locally. The most common controlled vocabularies in use for most museum collections include the Getty's Art & Architecture Thesaurus® (AAT),<sup>1</sup> the Union List of Artist Names® (ULAN),<sup>2</sup> and the Thesaurus of Geographic Names™ (TGN).<sup>3</sup> Non-Getty terminology sources include the Library of Congress' Thesaurus of Graphic Materials (LCTGM I and II),<sup>4</sup> The Revised Nomenclature for Museum Cataloging,<sup>5</sup> and ICONCLASS.<sup>6</sup> Each source varies widely in scope, content, and purpose, and no one source is sufficient within itself to address the wide variety of data required to adequately catalog museum collections.

Vocabularies and corresponding browser tools (most commonly the AAT) bundled with commercial management software provide users the capability to link appropriate descriptive terms to a record in the database. Not only does linking from an

---

<sup>1</sup>The Art & Architecture Thesaurus Browser, J. Paul Getty Trust, ©2000, (2 July 2001). <<http://www.getty.edu/research/tools/vocabulary/aat/>>.

<sup>2</sup>The Union List of Artist Names Browser, J. Paul Getty Trust, ©2000, (2 July 2001). <<http://www.getty.edu/research/tools/vocabulary/ulan/>>.

<sup>3</sup>The Thesaurus of Geographic Names, J. Paul Getty Trust, ©2000, (2 July 2001). <<http://www.getty.edu/research/tools/vocabulary/tgn/>>.

<sup>4</sup>The Library of Congress Thesaurus of Graphic Materials, 3 Nov. 2000, (2 July 2001). <<http://www.loc.gov/rr/print/tgm1/>>,<<http://lcweb.loc.gov/rr/print/tgm2/>>.

<sup>5</sup>James R. Blackaby, *The Revised Nomenclature for Museum Cataloging : A Revised and Expanded Version of Robert G. Chenall's System for Classifying Man-Made Objects* (Walnut Creek, CA: AltaMira Press, 1998).

<sup>6</sup>ICONCLASS, March 2001, (2 July 2001) <<http://www.iconclass.nl/>>

established source help encourage consistent use of terminology, but in the case of faceted or hierarchical thesauri, it enhances searching capabilities as well. Once linkages are established, queries can be broadened or narrowed based upon the position of the term in the hierarchy, synonyms, and/or related concepts. For example, a query on the term *portrait* may be much too broad to satisfy an information need, but it can be further narrowed to terms such as *self-portrait*, *group-portrait*, etc. Conversely, a query on *self-portraits* can be broadened to include all portraits indexed within the database. Moreover, a query on the term *portrait* will return items indexed under the term *portraiture* and vice versa.

Controlled vocabularies for museum use continue to gain attention in the literature as they improve in scope and accuracy and as more institutions migrate their data into electronic collection management systems. Several successful demonstration projects involving networked image and data resources have also raised the level of awareness among museum professionals. The Getty Research Institute (GRI) has been at the forefront of promoting controlled vocabulary usage and educating users on how to implement vocabularies into cataloging practices in order to add value to their intellectual property. In spite of efforts by GRI and others to disseminate the benefits of terminology control, the perception among most in the museum community is that vocabulary control is not being widely adopted.

To that end, this study was designed to assess the accuracy of that perception – to discover where museums stand in the divide between theory and practice when it comes to using controlled vocabularies to catalog their collections. Using a representative sample of thirty fine arts museums throughout the United States, each with an electronic

collections management system, the study seeks to answer the following research question: what factors affect the adoption and usage of vocabularies in art museum collection management databases?

## Literature Review

### I. Controlled Vocabularies

In the seminal publication, *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information*, controlled vocabularies are defined as “collections of words and phrases (called terminology) that are structured to show relationships between terms and concepts.”<sup>7</sup> As such, vocabularies function in two key capacities. First, they often represent agreement within a professional community as to preferred and consistent ways to describe and classify objects, behaviors, or characteristics. This function is also known as a “knowledge base,” or a body of knowledge within a subject area that documents usage, meaning and disambiguation.<sup>8</sup> Secondly, vocabularies can be used in electronic environments to facilitate search and retrieval by creating linkages between terms and ensuring standard usage. The majority of the literature involving controlled vocabulary usage focuses on their role in the retrieval context.

The absence of a controlled vocabulary in a catalog or database results in natural language indexing, or a list of the terms that appear somewhere within each record. To retrieve information, users must enter keywords that describe the concepts they are interested in without knowing if they are used in the database. The information and

---

<sup>7</sup> Elisa Lanzi, *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information* (The Getty Information Institute, 1998), 21. Also available online at:

<http://www.getty.edu/research/institute/vocabulary/introvocabs/>

<sup>8</sup> Ibid.

library science literatures are rife with studies comparing the relative precision and recall of controlled vocabulary versus free text searching.<sup>9</sup> Miller<sup>10</sup> summarizes some of the practical problems of free-text retrieval:

- 1) it is useful only when users want super-new information or super-narrow terms that are not yet incorporated into a vocabulary;
- 2) it leads to errors which are the corollaries of lexical and terminological peculiarities (synonymy, spelling variants, etc.);
- 3) ‘trial and error’ cannot be fruitful because every database is constructed differently and requires a search strategy suitable to the specific database;
- 4) users’ abilities to be proficient in searching depend almost entirely on the level of instructions worked out for the database.

A typical collection management database built upon a relational structure has (or should have) a documented data dictionary, table schema, and standardized instructions for query construction. Records typically do not represent “super new” or “super narrow” concepts although contemporary art collections might pose such a challenge. Applying Miller’s observations, it is possible these known deficiencies bolster the arguments for the success of controlled vocabularies in collection management systems. In writing about applying the ICONCLASS thesaurus to pictorial database systems, Berg states, “the recall and precision of queries will be considerably higher than it would be with

---

<sup>9</sup> See for example, Elaine Svenonius, “Unanswered Questions in the Design of Controlled Vocabularies.” *JASIS*, 37(5):331-340, 1986 and M.R. Muddamalle, “Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study.” *JASIS*, 49(10):881-887, 1998. Both offer comprehensive literature reviews in addition to exploration of specific research questions relating to indexing comparisons.

<sup>10</sup> Uri Miller, “Thesaurus Construction: Problems and Their Roots.” *Information Processing and Management*, 33(4):482, 1997.

uncontrolled keywords or free text, since ambiguity has been banned at the cost of ‘freedom of prose description.’”<sup>11</sup>

Terminology control is not without its drawbacks, however. In spite of the fact that most vocabulary sources are freely available online, museums must still make a substantial investment in personnel time and training in order to implement vocabulary control properly and efficiently. In contrast to libraries where collection cataloging is carried out by professionals who have been trained in a relatively uniform manner, museum cataloging is often completed by a range of staff, interns, volunteers, or docents with varying levels of academic and professional credentials. Frequently these individuals have a wide range of job duties in addition to managing collections information that can make using controlled vocabularies time-prohibitive.

While most vendors offer rudimentary instruction at the time the system is installed, organized training efforts such as professional seminars or academic courses are infrequent and often require more time and/or money than an institution or individual is willing to commit. Manuals and documentation are generally sparse, and it is often the case that catalogers decide to write their own documentation to better reflect the vocabularies’ use within a particular institution’s information system. Moreover, even when vocabularies are used and staff are properly trained, there are no guarantees that the terms chosen will be 100% accurate and consistent. “More workshops at national conferences, particularly in the use of AAT and ICONCLASS would help raise the level

---

<sup>11</sup> Jörgen van den Berg, “Subject Retrieval in Pictorial Information Systems.” 1995, (17 April 2001). <<http://www.iconclass.nl/texts/montreal.htm>>.

of image indexing and increase consistency of descriptive cataloging in image databases.”<sup>12</sup>

Finally, there are the vocabularies themselves. No one thesaurus or list of subject headings can be expected to fully represent the profound scope of knowledge required by any field, let alone one that is as broad and interdisciplinary as the history of art. Vocabularies as entities are slow to evolve and often do not reflect most recent language, making usage problematic for institutions and collections that deal with contemporary subject matter. In the case of the AAT and others, candidate terms must demonstrate literary warrant, or what is generally considered a substantial level of scholarly recognition and acceptance. Evaluating terms for warrant and relevance can be a subjective and lengthy process, and final decisions rarely please all members of the scholarly community. Benedetti renewed the debate over the scope and quality of folk art terms in the AAT, a pertinent example of disagreements over language continuing to rage nearly two decades since they were first noted. In her conclusion, Benedetti aptly observes, “words are powerful because they reflect people’s relative positions of power, their points of view, their ideas, and their feelings.”<sup>13</sup>

## **II. Controlled Vocabularies in the Museum Context**

As stated above, there are several vocabulary sources available to the museum community, in particular for fine arts and visual resources collections. The Getty Art History Information Program (AHIP), now known as the Getty Research Institute (GRI),

---

<sup>12</sup> Linda McRae, “Indexing Images for Subject Access: Controlled Vocabularies in the VISION Project.” *Art Documentation*, 19(2):2000, 8.

<sup>13</sup> Joan M. Benedetti, “Words, Words, Words: Folk Art Terminology – Why It (Still) Matters.” *Art Documentation* 19(1):2000, 19.

has been the primary impetus in the field of establishing vocabularies for cataloging and indexing works of art. The collected efforts of the Getty Vocabulary Program,<sup>14</sup> a subset of the GRI, have resulted in the CDWA (Categories for the Description of Works of Art) and three key vocabularies: AAT, ULAN, and TGN, that respectively target the ‘what’, ‘who’, and ‘where’ attributes of an object. As Patricia Harpring, Managing Editor of the Getty Vocabulary Program writes, “of all the information in a catalog record for an art object, the fields that record the names of people, places, and things are the most obvious targets where vocabularies could aid in retrieval.”<sup>15</sup> Getty vocabularies are intended for use in information systems, libraries, and museum cataloging systems for both text and image collections, and are updated regularly from internal and external contributions at no cost to users. The AAT, begun in the early 1980’s, is actively growing, and to date includes about 125,000 terms.

Harpring points to some possible retrieval complications that arise with inclusion of Getty vocabularies into collection management systems. For example, some terms can appear in all three vocabularies, resulting in general keyword searches that return indiscriminate results.<sup>16,17</sup> In these instances Harpring says “the most obvious way to improve results is to limit the search to a single common field . . .”<sup>18</sup> Other difficulties include homograph usage (i.e. *stretcher* as framing equipment, masonry unit, or furniture

---

<sup>14</sup> The Getty Vocabulary Program, J. Paul Getty Trust, ©2000, (22 June 2001). <<http://www.getty.edu/research/institute/vocabulary.html>>.

<sup>15</sup> Patricia Harpring, “Resistance is Futile: Inaccessible Networked Information Made Accessible Using the Getty Vocabularies.” *ASIS Annual Conference Proceedings*:1999, 838.

<sup>16</sup> Patricia Harpring, “How Forcible are Right Words! Overview of Applications and Interfaces Incorporating the Getty Vocabularies.” *Museums and the Web 1999 Conference Proceedings*, 18 March 1999, (11 Dec. 2000). <<http://www.archimuse.com/mw99/papers/harpring/harpring.html>>.

<sup>17</sup> Here, Harpring uses the example of painter – in AAT, it is an occupation, in ULAN it can be an artist’s name, and in TGN, it can be the name of a town.

<sup>18</sup> Harpring, *Museums and the Web*.

component.)<sup>19</sup> Homographs prove quite manageable in controlled, local environments, such as a museum's intranet; however, because each instance of the homograph occurs in separate locations in the hierarchy, and is given a unique numeric identifier, a cataloger can then place the identifier in the object record to create a relational link and thereby provide extremely accurate retrieval.

The field, no matter how dominated it seems by Getty products, is not completely without alternatives. In a 1993 article, Greenberg analyzes the differences in structure and usefulness between the then-incomplete AAT and the Library of Congress Thesaurus for Graphic Materials (LCTGM).<sup>20</sup> According to Greenberg, the AAT has a faceted structure that is geared toward a more specialized audience, while the LCTGM is categorized by hierarchical subdivisions that operate on a far more generalized level. This suggests that a user population of museum practitioners with more specialized knowledge than the general public would have greater success navigating within the AAT and using it as a cataloging tool. In 1995, two years after Greenberg's article, the LCTGM split into two separate databases: subject terms (TGM I), and genre and physical characteristic terms (TGM II).

One further example of a sophisticated tool for visual art classification is ICONCLASS, originally published between 1973 and 1985 in 17 printed volumes. Now available online, ICONCLASS offers "a classifying thesaurus that represents iconographic knowledge, arranged according to hierarchical principles. ICONCLASS

---

<sup>19</sup> Harpring, *ASIS*, 840.

<sup>20</sup> Jane Greenberg, "Intellectual control of visual archives: a comparison between the Art and Architecture Thesaurus and the Library of Congress Thesaurus for Graphic Materials." *Cataloging and Classification Quarterly*, 16(1):85-117, 1993.

contains at present about 24,000 definitions of objects, persons, events, situations and abstract ideas.”<sup>21</sup>

Spencer, in comparing ICONCLASS and the AAT for use in image collections, found the AAT to be more flexible than ICONCLASS for defining images because of the rigid, interlocking code structure of ICONCLASS. “Catalogers cannot always say what they want to say using these codes.”<sup>22</sup> Furthermore, she highlights that ICONCLASS was developed to deal with the traditional iconography of art history. As “traditional” is subjective and iconography is in constant flux, ICONCLASS may not have a place in collection management systems that catalog collections of contemporary, conceptual, or non-Western materials.

### **III. Controlled Vocabularies in Use: Pilot Projects**

In 1986, Judith L. Schulman of the Detroit Institute of Arts (DIA) published a detailed summary of a pilot electronic collections management project known as DARIS: The Detroit Art Registration Information System.<sup>23</sup> In her account, she highlights the ongoing and cooperative efforts of curators, educators, registrars and administrators in making refinements to both the system’s design and content. With regard to terminology control she writes:

Decisions on vocabulary access fields were difficult to make and once decided upon were changed as new and different types of objects were accessioned and catalogued. Standard vocabulary should never be construed as a limited

---

<sup>21</sup> Berg.

<sup>22</sup> Karen Spencer, “Authority and Vocabulary Control in Image Collections.” 1995, (27 October 2000). <<http://sunsite.berkeley.edu/Imaging/Databases/Fall95/papers/kspencer.html>>.

<sup>23</sup> Judith L. Schulman, “The Detroit Art Registration Information System (DARIS).” in *Museum Documentation Systems: Developments and Applications*, ed. R.B. Light, et al. (London: Butterworth & Co., Ltd., 1986), 80.

vocabulary. It requires thoughtful control and application of language. It cannot be emphasized enough that vocabulary control is the only guarantee for retrieval of like works.<sup>24</sup>

More recently the majority of research concerning controlled vocabulary implementation with museum information has centered on work done on networked resources, not stand-alone collection management systems as is the case of DARIS and this present study. In the mid-1990's a pilot project called 'a.k.a.' (standing for "Also Known As"), "experimented with strategies for gathering and manipulating terms from Getty vocabularies, and using the terms to broaden or narrow searches across [Getty] databases on the Web."<sup>25</sup> Harpring reports that end-users, generally accustomed to launching a search with a single search term or phrase were confused when confronted with a vocabulary from which to choose search terms. When search term broadening was performed automatically, it resulted in overly broad and imprecise results. However, "once users mastered the idea of interacting with the vocabularies . . . they were able to refine queries and were generally happier with the results."<sup>26</sup>

Another well-known example is AMICO, the Art Museum Image Consortium, started in October 1997 as a program of the Association of Art Museum Directors Educational Foundation, Inc.<sup>27</sup> One year later it incorporated separately under the leadership of *Archives and Museum Informatics* consultants David Bearman and Jennifer Trant. AMICO's online library holdings are currently estimated at 75,000 works of art plus their supporting documentation, all of which have been contributed by member institutions. While certain fields in the AMICO data dictionary indicate that they are

---

<sup>24</sup> Ibid., 80.

<sup>25</sup> Harpring, *ASIS*, 844.

<sup>26</sup> Ibid., 845.

<sup>27</sup> The Art Museum Image Consortium, (2 July 2001). <<http://www.amico.org>>.

AAT, ULAN or TGN controlled,<sup>28</sup> the overall quality of the data is largely subject to the wide-ranging cataloging standards of the contributing institutions. As a result, query results are often unpredictable and/or unsuccessful; for example, a search on “Van Gogh” retrieves only a fraction of the works in the database that are attributed to Vincent Van Gogh because many contributors have indexed his works under “Gogh.”

Inspired by an earlier project known as MESL (The Museum Educational Site Licensing Project), two additional art information sharing projects formed in the late 1990's: REACH (Record Export for Art and Cultural Heritage) and VISION (Visual Resources Sharing Information Online Network). Characterized as “twins of different mothers,”<sup>29</sup> REACH was born of the Getty Information Institute and the Research Libraries Group (RLG) in March 1997 while VISION was started a few months later by the Visual Resources Association (VRA) with support from both RLG and the Getty. Both were designed as test beds for interoperability standards including, among others, standardized vocabulary control through the use of thesauri.

The objective of REACH was to capitalize on the heterogeneity of data contained in museum collection management systems. In collaboration with software vendors, the goal was to devise an automated export tool that would output a common set of data elements considered necessary for basic access to cultural heritage materials. By using metadata crosswalks, these elements could then be linked to other common data structures including the Dublin Core and MARC catalog records for general research use in a networked environment. Vocabulary control was not central to this project,

---

<sup>28</sup> AMICO Data Specification, v.1.2, 4 June 2001, (2 July 2001).  
<<http://www.amico.org/AMICOLibrary/AMICO.dd.9902.html>>.

<sup>29</sup> Elisa Lanzi, “The REACH and VISION Projects: Improving Access to Art Information.” *Art Documentation*, 17(1):1998, 15.

although initially it was thought that a combined REACH/VISION database would share a web-based search engine with built-in vocabulary tools to facilitate searching.

In spite of the successful creation of a REACH element set,<sup>30</sup> the project terminated in November 1998 before it was complete due to a number of complications. The REACH Project Summary Report<sup>31</sup> cites problems extracting data from disparate and/or homegrown database systems and requests for export tools for other element sets competing for vendor's attention. Additionally, many of the eleven original REACH members contributed data to other projects such as AMICO or RLIN.

VISION, by contrast, had somewhat better success. Its objective was to provide a shared database of image records based upon contributions from thirty museum and library visual resources departments located throughout the United States.<sup>32</sup> A total of 1,667 records were entered into the database via a web-based template that provided access to nine recommended vocabularies including AAT, TGN, ULAN, ICONCLASS, Revised Nomenclature, LCTGM I and II, and LC's Subject and Name Authority files. Participants were strongly encouraged to use vocabulary sources, particularly for subject descriptors.

A subsequent study by McRae conducted on the VISION database compared a subset of building-related records and painting-related records with regard to the source of selected terms in two data fields: work type and subject.<sup>33</sup> Out of the nine available vocabularies, she found the AAT generally to be the most used for both object types, both for describing the work type (92.6% for buildings, 89% for paintings) and subject

---

<sup>30</sup> RLG REACH Element Set, 2 Sept. 1998, (2 July 2001). <<http://www.rlg.org/reach.elements.html>>.

<sup>31</sup> REACH Project Summary Report, 20 Jan. 1999, (2 July 2001). <<http://www.rlg.org/reach.html>>.

<sup>32</sup> Lanzi, *Art Documentation*, 16.

<sup>33</sup> McRae, 5. This study also looked at similar comparisons involving indexing depth, indexing specificity, and subject classes.

terms (73% for buildings, 30% for paintings). Local terms, or those not found in a standard vocabulary, were used in 40% of paintings records for subject terms. From this analysis McRae observed that when indexers opted to use controlled vocabularies they used only the tools with which they were most familiar, and even then to varying degrees of success and accuracy. “This study indicates that catalogers still need training and experience in using controlled vocabularies including the AAT. The tendency not to use some thesauri such as ICONCLASS and TGM suggests a general lack of familiarity with these thesauri.”<sup>34</sup> In summarizing the lessons learned from the REACH approach, she notes, “Even if shared cataloging is not a goal, successful retrieval is, whether from a stand-alone database or a shared database. Optimum retrieval depends upon record consistency that will not be achieved with a set of data elements alone.”<sup>35</sup>

These pilot projects, aimed at providing networked access to cultural heritage information, highlight the growing understanding of data standards and the roles they play in both local and networked information environments.

#### **IV. Controlled Vocabularies and the Networked Future**

In Chenhall’s 1975 book, *Museum Cataloging in the Computer Age*, he portends, “. . . Ultimately all museums will be linked together in some kind of network of computers so that any institution will have access to what has been recorded anywhere in the network.”<sup>36</sup> More than 25 years later, as the previous examples demonstrate, this is undoubtedly on its way to becoming a reality. At the core of these endeavors is the World Wide Web, allowing for institutions to provide the information they have

---

<sup>34</sup> McRae, 8.

<sup>35</sup> McRae, 8.

<sup>36</sup> Chenhall, 246.

compiled in their collection management systems via their websites in addition to contributing data to larger-scale initiatives. David Green, the Executive Director of National Initiative for a Networked Cultural Heritage (NINCH) makes a point of promoting the benefits to cultural heritage materials afforded by a networked society:

Digital networking has the immense potential of bringing wide, equitable access to the texts, objects, sounds, and sights that form our global cultural heritage. It has the added potential of being able to show the complex interrelationships among these objects, their histories, and their contexts. By networking the materials held by thousands of collectors and repositories, we can tell the stories of the objects of creation and their creators in a much richer and more rewarding way than ever before.<sup>37</sup>

There are numerous large-scale consortial efforts currently underway that exploit the benefits of both the Internet and increasingly more well defined data standards.

Many of these are aimed at creating electronic repositories of multi-format collections.

One example is Conceptual and Intermedia Online (CIAO),<sup>38</sup> a project started in 1997 to explore the possibilities of using Encoded Archival Description (EAD) to capture the complex and non-traditional nature of conceptual and digital art. Because conceptual art often eludes conventional cataloging practices and vocabulary tools often do not include adequate terminology, the application of EAD for organizing these complex collections into meaningful structures for search and retrieval is promising, though untested.

Another endeavor also utilizing EAD for union cataloging is Museums and the Online Archive of California (MOAC).<sup>39</sup> Since 1998 MOAC has been integrating museum collections information with information from libraries, special collections, and

---

<sup>37</sup> Lanzi, *Introduction to Vocabularies*, 1.

<sup>38</sup> Conceptual and Intermedia Online (CIAO), (1 July 2001). <<http://www.bampfa.berkeley.edu/ciao>>.

<sup>39</sup> Museums and the Online Archive of California (MOAC), (1 July 2001). <<http://www.bampfa.berkeley.edu/moac>>.

archival repositories. Richard Rinehart of University of California, Berkeley advocates exploring the possibilities of applying EAD and similar standards to museum collections because it would allow immediate access to museum data in integrated systems that include both archival and library materials.<sup>40</sup> Not only does EAD have the flexibility to allow collection-level through item-level description in one structure, but it also has tremendous scalability to support large-scale consortial efforts. At present it is too soon to know if EAD-based collaborations will have broad success or if controlled vocabularies used within the markup will continue to play an important part in resource discovery in these large-scale, cross-community endeavors.

*Artefacts Canada*, an initiative of the Canadian Heritage Information Network (CHIN), has been collecting museum object metadata for over 25 years, resulting in more than 2 million searchable object records.<sup>41</sup> To solve the problem of disparate or non-existent vocabulary standards in contributed data, the AAT has been integrated into the website's search engine with moderate success.<sup>42</sup> CHIN is currently developing a model for collection-level description that will further exploit the role of vocabularies in an attempt to "fill the semantic gap" between collection- and object-level metadata.<sup>43</sup>

## V. Controlled Vocabulary Usage – Previous Inquiries

Before addressing the methodology and results of this study, it is worthwhile to revisit two previous efforts to assess the state of controlled vocabulary usage in cultural

---

<sup>40</sup> Richard Rinehart, "Access to Art Collections Using Encoded Archival Description and Beyond: The Future of Large-Scale Consortium Projects." *Art Libraries Journal*, 26(3):33-39, 2001.

<sup>41</sup> *Artefacts Canada*, CHIN, ©2001, (5 July 2001). <[http://www.chin.gc.ca/Artefacts/e\\_artefacts\\_canada.html](http://www.chin.gc.ca/Artefacts/e_artefacts_canada.html)>.

<sup>42</sup> Heather Dunn, "Collection Level Description - the Museum Perspective," *D-Lib Magazine* 6(9):2000. (5 July 2001). <<http://www.dlib.org/dlib/september00/dunn/09dunn.html>>.

<sup>43</sup> Ibid.

heritage information systems. While neither addresses the same population nor uses the same approach as this study, they each provide a perspective into the attitudes and usage statistics of the museum community just five years ago.

The first is the aforementioned article by Spencer, who in 1995 informally spoke with librarians and curators of visual image collections to gauge the nature and extent of vocabulary usage.<sup>44</sup> She found that not having electronic versions of tools such as the AAT and ULAN was a barrier to use, and that many users were planning to consult them when they became offered electronically. Many users also admitted to being inconsistent with use of vocabularies and authority terms even when they were available, often times basing cataloging decisions on “historical knowledge” of the collection or training they received from predecessors.<sup>45</sup> Spencer observes a tendency for second and third generation catalogers to deviate from their predecessor’s methods, creating inconsistencies that only become apparent when collections are converted into electronic formats.

The second study is a 1996 survey of terminology in UK museums conducted by the Museum Documentation Association (MDA).<sup>46</sup> Of 2000 questionnaires mailed to museums of all types of collections throughout the UK, only 257 were completed. In spite of the low response percentage, the results indicate that “many of those institutions which reported the largest number of machine-readable records use – or are developing – thesauri.”<sup>47</sup>

---

<sup>44</sup> Spencer.

<sup>45</sup> Ibid.

<sup>46</sup> The Museum Documentation Association, MDA ©2001, (3 July 2001). <<http://www.mda.org.uk/>>.

<sup>47</sup> Survey of Terminology in [UK] Museums, 1996, MDA ©1996, (27 June 2001). <<http://www.mdocassn.demon.co.uk/survey.htm>>.

## Methodology

In this attempt to discover factors that affect adoption and usage of controlled vocabularies in art museum collection management databases, an interview protocol was designed for the basis of all inquiry (see Appendix A). The nature of questions ranges from simple reporting of known attribute information to those that solicit opinion and personal experience concerning terminology control. By incorporating a wide range of question types (attribute, beliefs, attitudes and behavior) the instrument design intentionally aims for both breadth and depth on the topics of vocabulary usage, museum resources, and involvement in networked initiatives.

Participants were selected according to a rigorous set of criteria to ensure that they were likely to be representative of current practice and behaviors, based on the premise that those museums that have vocabulary tools available are more likely to consider using them. Thus, all participants had to be currently using an electronic collections management system from a commercial vendor whose product includes the option to purchase and install vocabulary tools. The preliminary step was to identify candidate database systems using the widely recognized collection management database reviews from both the CHIN<sup>48</sup> and the Museum Documentation Association.<sup>49</sup> Eligibility was based upon the presence of a vocabulary browser or utility incorporated with the system for both data entry and query construction. From these candidates, the three receiving the highest review scores and those with the most comprehensive client lists were chosen. In order to achieve the desired diversity in the sample, it was

---

<sup>48</sup> Canadian Heritage Information Network, *Collections Management Software Review: Comparative Analysis*, 3<sup>rd</sup> ed., Canada: 2000.

<sup>49</sup> MDA Software Survey, MDA ©2000, (27 June 2001). <<http://www.mda.org.uk/software.htm>>.

important that the client museums represented a wide range of collection size, staff size, and geographic location. Additionally, only fine art collections in the United States were considered in order to maintain both consistency among the sample and to better conform to the body of previous and on-going research in this area.

Next, within each vendor's client list, candidates were further limited to institutions that are accredited by the American Association of Museums. Recognition by this professional body ensures that these museums have undergone rigorous peer review and adhere to the standards and best practices of the professional community.<sup>50</sup> AAM accreditation was utilized as both a criterion to further limit the sample and to identify those most likely to have high interest and awareness of issues involving data standards. Finally, ten institutions from each of the three vendors were selected based upon their relative collection size, staff size, annual attendance and geographic location.<sup>51</sup> Participating institutions and the demographic breakdown of the sample are included in appendices B and C.

The study questions were administered to the thirty individuals self-identified as the in-house contact responsible for or most familiar with the institution's database system. Very often these individuals were within the museums registrar's office, but in a few instances they were curators, catalogers, information technology professionals, or database administrators. Participants were assured that neither their identity nor their institutional affiliation would be compromised if they consented to participate. Twenty-

---

<sup>50</sup> AAM Services: Accreditation, 20 April 2001, (11 March 2001). <<http://www.aam-us.org/accredproginfo.htm>>.

<sup>51</sup> Collection size was estimated from institution websites and confirmed by participants self-reporting during participation. Staff size and attendance was obtained from *The Official Museum Directory*. Washington, DC: The American Association of Museums, 2001. Geographic distribution was determined in accordance with the regional classifications used by the American Association of Museums located at <<http://www.aam-us.org/aviso/map.htm>>.

two of the thirty agreed to be interviewed by telephone while 8 opted to respond to the questions via email. Not every participant provided answers to every question.

## **Results**

### **I. Vocabulary Usage**

Are American fine art museums in fact utilizing controlled vocabularies? Just under half of the sample (46.7%) said that the availability of a controlled vocabulary and browser utility was a factor in the decision to purchase to their software. Interestingly, only 50% of those institutions for which it was a factor actually have it installed. For the entire sample, 19 out of 30, or 63.3% have a vocabulary and browser utility installed as part of their collections management system, though only 11 of the 19 (57.9%; 36.7% of the total sample) are using it when entering data into the system. A total of 18 institutions among those who responded (60%) are using at least one controlled vocabulary reference when entering data into their systems.

The number of institutions who reported using a controlled vocabulary for data entry was spread comparably among the three collection management software packages. AAT and ULAN users each comprised half the clients for two of the systems and 2/5 for the third. Several participants, however, remarked that even when given the option of installing AAT or ULAN locally with their database system, they opted not to. Reasons included lack of ready access to the most recent term list updates without having to wait for system upgrades, reserving valuable disk space for their own data storage, and ease of use. For some, it remains easier to have a WWW browser open in the background than to have to travel through a large number of screens within the

database. The consistent distribution of vocabulary use within each vendor group along with the 57.9% use rate among those who have a vocabulary installed suggest that average use of controlled vocabularies, whether local or via the web, is in the 50-60% range.

In interviews, most participants commented on the circumstances that have led to their museum's current state of terminology control. Among the most representative responses on both sides of the argument are:

- Our curators are still emotionally attached to the old terminology, and as long as they stay consistent we don't mind.
- The newer curators are a lot more agreeable to adhering to some strict sense of how to catalog something whereas the ones who have been here forever don't buy in as quickly.
- We got into the habit of making up terms all over the place. Nobody sat down to say before you create your own terms why not look it up and see if it's already there.
- Vocabulary controls were not added to the database until one of the upgrades that were available sometime after it was purchased. By then, a system or pattern of entering data and new records was already in place.
- Using them [vocabularies] is tough because they're still in the process of building them, plus they are proprietary and people don't want to give up their terms to use somebody else's.
- Standardized terminology is a must.
- Vocabulary control is very important to us.

Fourteen respondents indicate regularly using the AAT, 14 report using ULAN, and 7 report using TGN. Nobody within the sample reported using ICONCLASS or LCTGM I or II, and only 2 respondents mentioned they occasionally consult The Revised Nomenclature. Eight museums use both AAT and ULAN, 7 use AAT and

TGN, and 7 use both ULAN and TGN. The summary of individual vocabulary usage is graphed below:

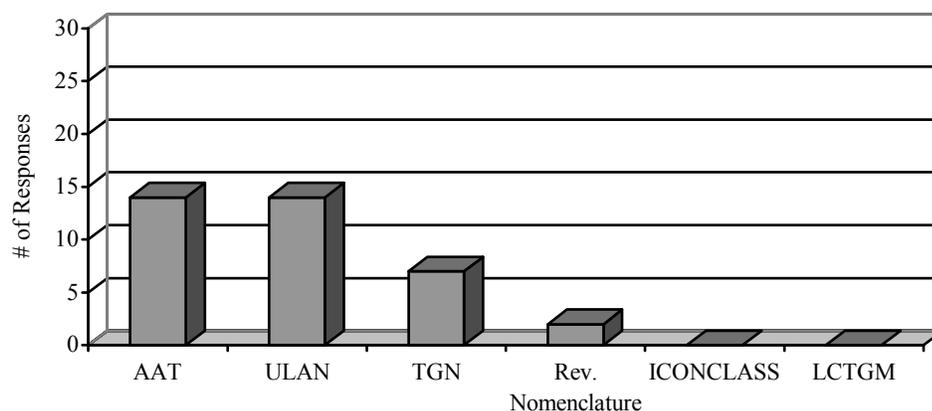


Figure 1: Summary of Vocabulary Usage

Comments garnered from interviews also reveal some of the most significant barriers to usage of individual vocabularies. AAT, the most widely integrated vocabulary, prompted the most feedback. The following statements represent the pertinent comments taken from interview responses in regard to factors affecting adoption behavior of specific vocabularies.

Regarding the AAT:

- The AAT is too complicated. It's over-structured. Most people feel it's very complicated . . . I envy the simplicity of the Library of Congress subject headings. Cataloging a work of art is far more complex than cataloging a book.
- I find it annoying. [The AAT] goes too far, it's too clunky. Also, our collection is encyclopedic – it covers all times and all parts of the world. I just don't know that using it is that helpful for our collection.
- Intellectually, it's still hard to grasp. People don't know how deep or broad to go into the hierarchy.

- We don't think of terms within a hierarchy - the curators don't see or use the terms in the same way.
- . . . unwieldy hierarchical string . . .
- Yes it's installed, but we just don't have the staff to deal with it.
- . . . steep learning curve . . .
- It's a good functionality and a great feature, we're just not using it. Part of the reason is that the terms for the types of objects we have are not highly developed. But the hierarchy is a good concept.
- It's useful once you get the hang of it.

#### Regarding ULAN:

- We tried to use ULAN as a name source, but not only is it geared toward Western artists, but it's odd with names in [other languages].
- For non-Western art, I don't even bother to look at those tools. An old French dictionary has more Japanese artists listed than ULAN.
- ULAN is very limited for Asian collections – well, all non-Western, and we have a broad collection. We'd like to only have to use one or two sources.
- There are very clear cases when ULAN is wrong, so our local information is preferred.
- ULAN comes in handy sometimes, but sometimes we don't trust the dates. But probably of all of them ULAN is the easiest to use.

#### Regarding TGN:

- TGN - doesn't have all places we need and preferred term isn't our preferred term - they use language of country rather than English. Sometimes we want to use historical terms for place rather than current term.
- I would use TGN also although I get too caught up in the information because it's too interesting, so I try not to go in there.

Statistical analysis of the data revealed strong correlations between two sets of variables: those who use the AAT and TGN ( $p < 0.05$ ) and those who use ULAN and TGN ( $p < 0.05$ ). The chi-square values, each also well below the accepted 0.05 cutoff, indicate there are significant relationships between variables. Unfortunately, due to the very small sample size it is problematic to conclude with any level of confidence that use of one Getty vocabulary significantly correlates to use of another. It is more plausible to assume that institutions that are familiar with how to use one or more of the Getty vocabularies are likely to consider using others as time and individual data requirements permit.

Table 1  
Statistical Analysis on Selected Vocabulary Variables

		AAT	ULAN
TGN	Pearson's R	.584	.584
	p	.001	.001
	Pearson's Chi-Square	.002	.002
	N (sample size)	29	29

Aside from an external vocabulary, 26 of the 30 institutions (86.7%) reported that they are using customized lists of terms (referred to throughout this study as authority control.) Most often these term lists are used in fields that describe object types, materials, and geographic origin, but also include artist/creator names, styles, or periods. Preferred sources for terms mentioned by interviewees include Getty tools, the Revised Nomenclature, Hall's *Dictionary Of Subjects & Symbols In Art* and the Bibliography of the History of Art (BHA). Other common sources indicated were curators, educators, scholarly publications, and institutional memory ("the way it's always been done"). One-third of the sample uses only their own authorities when

cataloging museum objects. Only two institutions utilize neither vocabulary nor authority control in their systems.

## II. Museum Resources

Just under half of all participant institutions reported three or fewer staff members responsible for entering new data into the database. The maximum number of cataloging staff in an institution using vocabulary control is 12, and the average is 4.1. The average number of cataloging staff among institutions not using vocabulary control is 8.9. These numbers suggest that fewer staff with data maintenance responsibility may be a factor in successful adoption of controlled vocabulary practices.

The average level of academic training among catalogers was difficult to ascertain with any accuracy, but typical responses were either bachelors or masters degrees in studio art, art history, or museum studies. In five instances, respondents indicated that they or someone on the cataloging staff held a library science degree. All five of these institutions use both authority control and vocabulary control.

This strong correlation comes as no surprise, as many museums are beginning to awaken to the advances that libraries have made in terminology control and data standards in cataloging systems. A recent article published by members of the Solomon R. Guggenheim Museum's documentation program points to this very trend: "Automated libraries are a generation ahead of collection management systems and librarians can share with their colleagues their hard-won knowledge gained from their experiences . . ." <sup>52</sup> In that vein, one interviewee observed, "libraries have developed

---

<sup>52</sup> Tasha Seren, et al. "Integrated Art Documentation: the Guggenheim Perspective" *Art Documentation*, 20(1):31-35, 2001. The Guggenheim Museum was not among the participants in this study.

more of a best practice system, but museums have systems that just automate museum work – which is just practice, not best practice.” Expressing doubt that the two disciplines can ever fully integrate, another participant contributed:

Library systems are rigid in categorization, but museums resist any hard and fast rules. We don't like to pigeon hole curators into trying to name things the way they should be. Instead, we use “guidelines,” and I don't know if there is ever going to be a merging.

Twenty-six of the 30 museums have at least 90% of their permanent collection cataloged in their present database system. The total number of electronic catalog records created by institutions is often much greater than the number of objects in the collection as many museums create records for objects brought in as loans or possible acquisitions. In relation to the 1996 MDA study that found museums with the largest number of records more likely to use or develop terminology sources, the results of this present study concur for the most part. Out of the five museums in the sample with the largest number of records (100,000 and greater), only one uses controlled vocabulary, but all five use authority control. Within the next tier of five, those with electronic holdings in the range between 40,000 and 99,999 records, four out of five are using vocabulary control, and all five are use authority control. The smallest five collections in the sample (0-5999 records) registered two out of five using vocabulary control, three out of five using authority control.

### **III. Attitudes and Behaviors Toward Collections Management**

In order to gauge whether user's attitudes toward and experiences with their collection management software influence vocabulary adoption behavior, participants

were asked to rate their satisfaction with several aspects of the system based upon their personal experiences. Responses were ratings on a Likert scale of 1 to 5, with the following designations: Highly Satisfied (5), Satisfied (4), Somewhat Satisfied (3), Somewhat Dissatisfied (2), Dissatisfied (1). The tabulated results of these rankings appear below.

Table 2  
Summarized Results of Satisfaction Indicators

	# responses (N=29)	Mean	Median	Mode
Satisfaction with the vocabulary browser tool included with the database (if applicable/installed)	17	3.65	4	4
Satisfaction with the quality of data contained in the database (in regard to consistency, thoroughness, accuracy, etc.)	29	3.55	4	4
Satisfaction with the data contained in query results (in regard to consistency, thoroughness, accuracy, etc.)	27	3.86	4	4
Overall satisfaction with the database for daily use to accomplish job objectives	27	4.25	5	5

Satisfaction with the browser tool revealed a strong correlation among institutions that use the AAT when cataloging records ( $R = .509$ ,  $p < .001$ ). As previously stated, the primary vocabulary being incorporated into commercial collection management software is the AAT; thus, users at institutions that are using the AAT have more experience using the tools included with the database. The fact that users have an overall positive view of the tools is promising news to both vendors and the museum community. Again, as the sample size is quite small, commenting on the significance of any correlations from the dataset is problematic.

Interestingly, satisfaction with data quality negatively correlates with the number of items in the collection ( $R = -.422, p < 0.05$ ), the total number of employees ( $R = -.436, p < 0.05$ ), and the museum's annual attendance ( $R = -.499, p < 0.01$ ). One assumes that the larger an institution's operations are, the more difficult it becomes to maintain a consistent standard of thorough and accurate data. These numbers align closely to the previous observations that the more staff members a museum has cataloging the data, the less likely it is to use vocabulary control. As noted above, only one of the institutions with 100,000 electronic records or more is using vocabulary control. The average satisfaction with data quality among those five largest institutions was below the sample average at 3.2, and among those responses was one of only two "dissatisfied" (1) ratings received for any of the questions measuring satisfaction.

Since the median length of current database ownership for the sample (excluding previous systems and older versions of the same software) is just three years and four months<sup>53</sup>, many institutions are still going through the time-intensive and tedious process of cleaning up records that have been converted out of older systems while also maintaining day-to-day operations. During interviews, many participants spoke in detail about how their institution's priorities concerning collections cataloging affected data. For the most part, implementing new data quality standards has taken a lower priority behind entering new object records and cleaning up records already in the system.

Representative comments from those interviewed include:

- A big chunk of my time right now is spent going back and cleaning up the data from previous conversions.

---

<sup>53</sup> Short-term system ownership reflects the rapid changes in the software and technology industries along with a surge in system turnover just prior to the year 2000 due to y2k incompatibility issues.

- Our museum has decided to go for breadth rather than depth in the system, so we're getting a lot of data in, just not as clean as one would like.
- The thought of going back to clean up is just overwhelming. Sitting at the computer and having to enter these terms is just so tedious and our staff have so many other things to do. We even had resistance to even computerizing data in the first place.

A number of museums indicated that they now have or are forming committees to address data standards including vocabulary control in the collections database, but these cooperative efforts among curators, registrars, educators, and IT staff are slow to develop and slow to reach consensus. In general, vocabulary control in collections management systems has yet to register as a significant priority for many institutions, but forming committees demonstrates that many are making an important first step.

Emphasis on speed over quality was a trend noted by many. Efforts to populate databases have been spurred on by the move to get collections data linked to the Web at a rapid pace. Remarks to that effect include, "We've been trying to finish the basic core fields in order to get something on the web" and "most major museums have a website now, and its competitive to get collections up there." One registrar from a museum whose collection is completely web-accessible noted, "I'm grateful for our administration that they've been so forward thinking. They've made the commitment." Another interviewee in the opposite position commented, "Getting our data in shape needs to be an institutional mandate of how to spend time, energy, and money, and it needs to come from the director."

The perception that most museums have their collections online however, may not be entirely accurate. From this sample only 9 museums of 29 responding (31%) say that the at least some of the information in their database is available through their

website (whether directly or through data export to an external system.) One-third of these museums are affiliated with universities that maintain some form of electronic union catalog of museum, library and archival holdings. Even when museums do make collections data available, it is typically only a percentage of the total collection or records that represent the institution's signature works. In the words of one database administrator, "what good is having it on the web unless the whole database is there so you get a complete search?"

As the third indicator of satisfaction, participants were asked whether or not they use any vocabulary tools when constructing queries within the collection management system. Eleven out of 30 (36.7 percent) responded that they did, though most did not make a clear distinction between using a utility that allows manipulation of the query based on term linkages and choosing a pre-defined term from a drop-down list of authority terms. The results of perceived user satisfaction with query results did not correlate to any other factors, including whether or not an institution uses controlled vocabularies. The recurring comment concerning query satisfaction was typically a variation of "what you get out is only as good as what you put in."

Some users expressed frustration at the limitations of the query tools provided with the software and constraints in the database design itself that prevent searching on some fields. Many noted that they frequently had to construct multiple queries to get a complete set of results. In spite of these frustrations however, satisfaction with query results received a slightly higher score than satisfaction with the quality of the data itself, most likely because participants recognize the limitations of the system and have learned ways to work around them.

The high score recorded for the final statement regarding overall satisfaction with the database system for daily use reflects a positive trend. Museum personnel are growing increasingly more comfortable with using technology, and the systems and tools they are using continue to improve. One participant, relatively new to her system remarked, "I'm happier with it every day." Institutions who were the least satisfied typically had older installations of their systems and were either awaiting or contemplating upgrades. In one instance, an interviewee felt the system his museum had purchased was simply not a good fit for the nature of the collection and the types of functions they needed the system to fulfill.

Though it does not reflect in the satisfaction rating, a clear division in attitudes emerged between those museums who see their databases purposed primarily as a registrar's tool, and those who consider it more of a global information system. "Collections management systems for internal use were not geared for making data pretty for the public. [It is] a workhorse tool for large numbers of objects with a large amount of information so it can work for you and leave a trail." Those in the latter category tended to express doubts and frustration with the tools currently available on the market for linking data from the database directly to the website and integrating large numbers of related images. While a few felt they had been promised something better than they purchased, most were optimistic that the tools will continue to improve with time and experience. All acknowledged that it is better to have something than nothing, and for overall daily use, the systems gained a high satisfaction rating.

#### **IV. Networked Initiatives**

Half of the institutions in the study indicate that they share their collections data in some form of networked access. Though not a factor in selection for participation in the study, AMICO members compose exactly 1/3 of the sample. Five out of the ten AMICO members are among those using controlled vocabularies. Several non-AMICO member institutions expressed interest or plans to contribute to the library in the future. Costs and time involved to prepare data for export were frequently cited as drawbacks or barriers to AMICO membership, but overall it was highly desired among those who are not currently members. From an administrative point of view however, the idea of contributing resources to a large-scale networked initiative is generally secondary to providing information via institutional websites. Once museums achieve an accepted level of data quality, they may revisit the decision to contribute to data sharing projects such as AMICO as a way of extending their internal efforts toward data quality and access.

Other types of partnerships mentioned by study participants include contributions to union library catalogs as previously stated, and established consortiums such as CIAO. Two museums indicated they are in the planning and pilot phases of data sharing projects with other museums or regional institutions.

A representative from one museum indicated that their institution's interest in networked access is shifting toward metatagging projects, standards initiatives and search engine tools for cross-database querying. "We're moving away from wanting to contribute physical datasets to other databases." Instead, the idea is for a sophisticated portal that links numerous distributed databases, searchable with one engine that has

built-in vocabulary tools to handle differences in terminology as well as spelling variants and synonyms. The portal concept would alleviate some of the costs and time required to export customized datasets, and it is, in theory, infinitely scalable. Developments in metadata crosswalks among disparate element sets will also help to increase the accuracy and precision of retrieved sets from distributed databases.<sup>54</sup> Controlled vocabularies, while still of great importance locally, may prove to be not as necessary in the networked realm when it comes to guaranteeing high precision retrieval capabilities.

## **Conclusion**

Assuming that just over half of the nation's fine art museums have adopted vocabulary control as a regular cataloging practice, the professional museum community still has a long way to go to achieve objectives toward this particular aspect of data standardization. End-users can no longer claim that they are waiting for electronic versions of the vocabularies or that they are blindly following in the footsteps of their predecessors. The advent of the Internet has made both excuses obsolete, for now access to both vocabularies and best practices for museum data standards is ubiquitous.

Additionally, the role that collections management software plays within museums will continue to evolve. Personnel in education, user services and exhibitions departments are finding ways to adapt the software for their unique purposes, and in doing so, challenge the traditional view that these databases are to be used as exclusively as registrar's tools. In 1997 Besser observed, “. . . recent advances [re: technological

---

<sup>54</sup> See for example Murtha Baca, ed. “Metadata Standards Crosswalks,” Introduction to Metadata: Pathways to Digital Information, J. Paul Getty Trust, © 2000, (17 June 2001).  
<[http://www.getty.edu/research/institute/standards/intrometadata/3\\_crosswalks/index.html](http://www.getty.edu/research/institute/standards/intrometadata/3_crosswalks/index.html)>

limitations] are likely to promote the convergence between these two camps.”<sup>55</sup> With increasing levels of use for more diverse purposes, museums may learn through experience how valuable a standardized vocabulary can be when it comes to wanting to capitalize on their data assets.

Murtha Baca, Head of Vocabulary Standards for the Getty Research Institute, cites knowledge – or lack thereof – chief among the obstacles to using vocabularies. “Vocabularies have to be available and museums have to know about them and know that they are important.”<sup>56</sup> She also points to the changing roles of museum personnel, noting that previously non-existent job titles such as “Data Standards Administrator” will start to become more prevalent, and with them, more comprehensive and enforceable data standards programs. Additionally, Baca indicated that a work on a new Getty publication is currently underway – one that addresses subject access for image collections.

The gap between awareness of vocabularies and skill in using them perhaps could be further narrowed if institutions commit to continuing professional education for staff that extends beyond introductory training seminars. Dissemination of innovation also could be greatly enhanced if collections management personnel were to develop local, regional, or national networks among themselves to foster support and collaboration.

No single profile accurately characterizes the type of museum that is a good candidate for successful implementation of vocabulary control standards, though the

---

<sup>55</sup> Howard Besser, “The Transformation of the Museum and the Way It’s Perceived.” In *The Wired Museum*, ed. Katherine Jones-Garmil. (Washington, DC: The American Association of Museums, 1997), 160.

<sup>56</sup> Murtha Baca, 21 June 2000, telephone interview by author.

results of this study strongly indicate that the presence of individuals with information and library science education is a factor in improving adoption behavior. While this study also shows that institutions with smaller numbers of records and fewer staff actively involved in cataloging are more likely to use vocabularies, actual implementation varies widely with regard to the nature of their collections, the priorities of their administrations, and individual staff workloads. Among those participants who are currently using vocabularies, many indicated that they are actively writing their own documentation and conducting small training seminars within their museums to supplement training received from vendors.

The results of this study show long-standing habits and institutional traditions emerging as the most significant barriers that need to be overcome before a museum can objectively approach terminology control. The advent of networked resources and the drive to the Internet have sparked a great deal of change in the way museums view their intellectual capital, but the realities of change remain governed by availability of resources including staff time, training, financial and administrative support. Close behind tradition are the tools themselves. Complaints of complexity, inaccuracy, steep learning curves and insufficient scope are serious concerns for many, and factors that have a profound affect on an institution's decision to adopt one or more vocabularies. Finally, the time involved presents a substantial obstacle to museum personnel who are barely managing to complete daily workloads without the added burden of a terminology control project.

## Appendix A: Interview Protocol

(For items marked with “\*” please estimate if exact figures are not known or readily available)

1. What software package is your institution currently using for collections management?  
What version?
2. Is this the first commercial database system your institution has purchased?
3. How long have you had this system? \_\_\_yr(s) \_\_\_mo.(s) \*
4. Total number of objects owned by the museum? \*
5. How many museum-owned objects are currently recorded in the database? \*  
(For instances of multiple objects with one accession number, estimate the number of accession numbers)
6. How many total object records are currently in the database? (accounting for instances where the museum records loans, potential acquisitions, etc.) \*
7. How many staff enter new data into the database? \*  
In general, what positions do these staff members hold?  
What is their level of academic/professional training?
8. Have you personally used any vocabulary tools in your own career or training?  
If yes, what tools, and describe your impressions or experiences.
9. Was the availability of vocabulary tools or a built-in vocabulary browser (such as the Getty’s Art and Architecture Thesaurus, Union List of Artist Names, Thesaurus of Geographic Names, ICONCLASS, or Library of Congress Thesaurus of Graphic Materials) a factor in your institution’s decision to purchase this system?
10. Are any vocabulary tools (such as those named above) installed on your system?  
If yes, which ones? continue to question 11.  
If no, why not? Skip to question 13.
11. Do catalogers at your institution regularly use formal vocabulary control sources when *entering* new records?  
If yes, which ones?
12. Do database users at your institution regularly use vocabulary tools when *querying* the database? (e.g. Querying on a broader or narrower term in a hierarchy)
13. Do catalogers at your institution regularly use authority control (i.e. an in-house or customized list of terms, not a formal thesaurus) when entering new records?

14. Does your institution offer staff training in how to use the database?

If yes, describe the nature and availability of training. In what aspects?  
(e.g. data entry, querying, report writing, world wide web access, etc.)

15. Please rank the following based on your personal experience, and feel free to comment on any of the selections:

Satisfaction with the vocabulary browser tool included with the database?  
(If applicable/installed)

Highly Satisfied	Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Dissatisfied
------------------	-----------	--------------------	-----------------------	--------------

Satisfaction with the quality of data contained in the database?  
(in regard to consistency, thoroughness, accuracy, etc.)

Highly Satisfied	Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Dissatisfied
------------------	-----------	--------------------	-----------------------	--------------

Satisfaction with the data contained in query results?  
(in regard to consistency, thoroughness, accuracy, etc.)

Highly Satisfied	Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Dissatisfied
------------------	-----------	--------------------	-----------------------	--------------

Overall satisfaction with the database for daily use to accomplish job objectives?

Highly Satisfied	Satisfied	Somewhat Satisfied	Somewhat Dissatisfied	Dissatisfied
------------------	-----------	--------------------	-----------------------	--------------

16. Is the database, whether in part or whole, accessible to the public?

If yes, how? (check all that apply):

- kiosks located in gallery space in conjunction with objects on view
- a designated computer area apart from the collection
- by appointment only
- via the museum's website
- via another website \_\_\_\_\_
- other \_\_\_\_\_

17. Does your institution currently collaborate or have future plans to collaborate in any form of networked access to your collections data with any other institution? (i.e. museum-library partnerships, AMICO membership, union catalogs or databases, etc.)

18. Do you have any other comments or experiences relating to your institution's use of controlled vocabularies or vocabulary tools that you would like to share?

**Appendix B: List of Participating Institutions**

Albright-Knox Art Gallery  
Birmingham Museum of Art  
Corning Museum of Glass  
Dallas Museum of Art  
Delaware Art Museum  
Denver Art Museum  
High Museum of Art  
Hood Museum of Art  
Huntington Museum of Art  
Indianapolis Museum of Art  
J. Paul Getty Museum  
Johnson Museum of Art, Cornell University  
Joslyn Art Museum  
Los Angeles County Museum of Art  
Metropolitan Museum of Art  
Minnesota Museum of American Art  
Museum of Contemporary Art, Chicago  
Museum of Fine Arts, Boston  
Museum of Fine Arts, Houston  
Nelson-Atkins Art Museum  
North Carolina Museum of Art  
Philbrook Museum of Art  
Phoenix Art Museum  
Seattle Art Museum  
Smith College Art Museum  
Sterling and Francine Clark Art Institute  
University of Wyoming Art Museum  
Walters Art Gallery  
Weatherspoon Art Gallery  
Yale University Art Gallery

### Appendix C: Demographic Breakdown of Sample

Geographic Region	No. of Employees	Annual Attendance	Collection
<b>Midatlantic</b>	41	72000	31000
	50	79155	11000
	90	288275	40000
	120	135524	5500
	140	300000	30000
	2530	4900000	3000000
Regional Mean	495.2	962492.3	519583.3
<b>Midwest</b>	107	273676	5700
	203	259339	50000
	403	307631	30000
Regional Mean	237.7	280215.3	28566.7
<b>New England</b>	30	38352	59500
	40	35003	24000
	54	100970	100000
	95	160000	4000
	796	1132705	350000
Regional Mean	203	293406	107500
<b>Plains</b>	9	40000	3500
	20	50000	7000
	106	248553	12000
	224	500000	20000
	235	586971	55000
	606	1500000	35000
Regional Mean	200	487587.3	22083.3
<b>Southeast</b>	24	28000	6000
	28	33072	9000
	63	105440	8500
	90	214849	21000
	140	552603	11000
	148	311355	3600
Regional Mean	82.2	207553.2	9850
<b>West</b>	104	567000	17000
	245	505264	22000
	296	80000	96000
	348	1547310	80000
Regional Mean	248.3	674893.5	53750
<b>Total Mean</b>	246.2	498434.9	138243.3

## Bibliography

- American Association of Museums Services: Accreditation. 20 April 2001.  
<<http://www.aam-us.org/accredproginfo.htm>> (11 March 2001).
- AMICO: The Art Museum Image Consortium. <<http://www.amico.org>> (2 July 2001).
- AMICO Data Specification, v. 1.2, 4 June 2001,  
<<http://www.amico.org/AMICOLibrary/AMICO.dd.9902.html>> (2 July 2001).
- The Art & Architecture Thesaurus Browser. J. Paul Getty Trust, ©2000.  
<<http://www.getty.edu/research/tools/vocabulary/aat/>> (2 July 2001).
- Artefacts Canada*, Canadian Heritage Information Network, ©2001.  
<[http://www.chin.gc.ca/Artefacts/e\\_artefacts\\_canada.html](http://www.chin.gc.ca/Artefacts/e_artefacts_canada.html)> (5 July 2001).
- Baca, Murtha, ed. "Metadata Standards Crosswalks," Introduction to Metadata: Pathways to Digital Information. J. Paul Getty Trust, ©2000.  
<[http://www.getty.edu/research/institute/standards/intrometadata/3\\_crosswalks/index.html](http://www.getty.edu/research/institute/standards/intrometadata/3_crosswalks/index.html)> (17 June 2001).
- . Telephone interview by author. 21 June 2001.
- Benedetti, Joan M. "Words, Words, Words: Folk Art Terminology – Why It (Still) Matters." *Art Documentation* 19(1):14-21, 2000.
- Berg, Jörgen van den. "Subject Retrieval in Pictorial Information Systems." Proceedings of the 18<sup>th</sup> International Congress of Historical Sciences, Round Table 34: Electronic Filing, Recording, and Communication of Visual Historical Data: Montreal, 1995.  
<<http://www.iconclass.nl/texts/montreal.htm>> (17 April 2001).
- Besser, Howard. "The Transformation of the Museum and the Way It's Perceived." In *The Wired Museum*, ed. Katherine Jones-Garmil. (Washington, DC: The American Association of Museums, 1997), 153-169.
- Blackaby, James R. *The Revised Nomenclature for Museum Cataloging : A Revised and Expanded Version of Robert G. Chenall's System for Classifying Man-Made Objects*. Walnut Creek, CA: AltaMira Press, 1998.
- Canadian Heritage Information Network. *Collections Management Software Review: Comparative Analysis*, 3<sup>rd</sup> ed., Canada: 2000.

Conceptual and Intermedia Online (CIAO). <<http://www.bampfa.berkeley.edu/ciao>> (1 July 2001).

Dunn, Heather. "Collection Level Description - the Museum Perspective." *D-Lib Magazine* 6(9):2000. <<http://www.dlib.org/dlib/september00/dunn/09dunn.html>> (5 July 2001).

The Getty Vocabulary Program. J. Paul Getty Trust, ©2000.  
<<http://www.getty.edu/research/institute/vocabulary.html>> (22 June 2001).

Greenberg, Jane. "Intellectual control of visual archives: a comparison between the Art and Architecture Thesaurus and the Library of Congress Thesaurus for Graphic Materials." *Cataloging and Classification Quarterly*, 16(1):85-117, 1993.

Harpring, Patricia. "Resistance is Futile: Inaccessible Networked Information Made Accessible Using the Getty Vocabularies." *ASIS Annual Conference Proceedings*: 833-847, 1999.

----- . "How Forcible are Right Words! Overview of Applications and Interfaces Incorporating the Getty Vocabularies." *Museums and the Web 1999 Conference Proceedings*, 18 March 1999.  
<<http://www.archimuse.com/mw99/papers/harpring/harpring.html>> (11 December 2000).

ICONCLASS. March 2001. <<http://www.iconclass.nl/>> (2 July 2001).

Lanzi, Elisa. *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information*. The Getty Information Institute, 1998. Online:  
<<http://www.getty.edu/research/institute/vocabulary/introvocabs/>>.

----- . "The REACH and VISION Projects: Improving Access to Art Information," *Art Documentation*, 17(1):15-18, 1998.

The Library of Congress Thesaurus of Graphic Materials I & II. 3 Nov. 2000.  
<<http://www.loc.gov/rr/print/tgm1/>> & <<http://lcweb.loc.gov/rr/print/tgm2/>> (2 July 2001).

McRae, Linda. "Indexing Images for Subject Access: Controlled Vocabularies in the VISION Project." *Art Documentation*, 19(2):4-9, 2000.

MDA Software Survey. MDA ©2000. <<http://www.mda.org.uk/software.htm>> (27 June 2001).

Miller, Uri. "Thesaurus Construction: Problems and Their Roots." *Information Processing and Management*, 33(4):481-493, 1997.

Muddamalle, M.R. "Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study." *JASIS*, 49(10):881-887, 1998.

The Museum Documentation Association. MDA, ©2001. <<http://www.mda.org.uk/>> (3 July 2001).

Museums and the Online Archive of California (MOAC).  
<<http://www.bampfa.berkeley.edu/moac>> (1 July 2001).

*The Official Museum Directory*. Washington, DC: The American Association of Museums, 2001.

REACH Project Summary Report. 20 Jan. 1999. <<http://www.rlg.org/reach.html>> (2 July 2001).

Rinehart, Richard. "Access to Art Collections Using Encoded Archival Description and Beyond: The Future of Large-Scale Consortium Projects." *Art Libraries Journal*, 26(3):33-39, 2001.

RLG REACH Element Set. 2 Sept. 1998. <<http://www.rlg.org/reach.elements.html>> (2 July 2001).

Schulman, Judith L. "The Detroit Art Registration Information System (DARIS)," in *Museum Documentation Systems: Developments and Applications*, ed. R.B. Light, et. al. (London: Butterworth & Co., Ltd., 1986), 77-88.

Seren, Tasha, Deirdre Donohue, and Lynn Ann Underwood. "Integrated Art Documentation: the Guggenheim Perspective." *Art Documentation*, 20(1):31-35, 2001.

Spencer, Karen. "Authority and Vocabulary Control in Image Collections." 1995, <<http://sunsite.berkeley.edu/Imaging/Databases/Fall95/papers/kspencer.html>> (27 October 2000).

Survey of Terminology in [UK] Museums, 1996. MDA, ©1996.  
<<http://www.mdocassn.demon.co.uk/survey.htm>> (27 June 2001).

Svenonius, Elaine. "Unanswered Questions in the Design of Controlled Vocabularies." *JASIS*, 37(5):331-340, 1986.

The Thesaurus of Geographic Names. J. Paul Getty Trust, ©2000.  
<<http://www.getty.edu/research/tools/vocabulary/tgn/>> (2 July 2001).

The Union List of Artist Names Browser. J. Paul Getty Trust, ©2000.  
<<http://www.getty.edu/research/tools/vocabulary/ulan/>> (2 July 2001).