

2008

Individual Efforts, Democratic Designs: Collaborative Tagging and the Future of Museum Cataloging

Daniel Brennan
Seton Hall University

Follow this and additional works at: <http://scholarship.shu.edu/dissertations>

 Part of the [Arts and Humanities Commons](#)

Recommended Citation

Brennan, Daniel, "Individual Efforts, Democratic Designs: Collaborative Tagging and the Future of Museum Cataloging" (2008).
Seton Hall University Dissertations and Theses (ETDs). 886.
<http://scholarship.shu.edu/dissertations/886>

**Individual Efforts, Democratic Designs:
Collaborative Tagging and the Future of Museum Cataloging**

Daniel Brennan

**Submitted in partial fulfillment of the requirements for the degree of Master of Arts
in Museum Professions
Seton Hall University
December 2007**

Jürgen Heinrichs, Ph.D., Faculty Advisor

Abstract

Current museum collections cataloging practices, many of which are borrowed from the library and archives fields, are ill suited for describing museum objects. Initiatives on the part of museums to digitize collections and make them accessible via the web have only served to exacerbate the fundamental flaws of museum cataloging. These flaws include an inability to accurately convey the contextual importance of museum objects, the perpetuation of a semantic gap which exists between museum professionals and patrons, and the failure of cataloging to keep pace with ongoing digitization projects. In broader terms, museum cataloging by conventional practices ensures that control over the description of collections remains entirely in the hands of the institution.

The advent of Web 2.0 portends to transform the traditional museum-patron paradigm. In particular the practice of collaborative tagging, a process in which individual users “tag” digital resources with descriptive terms of their own choosing, holds the potential to remedy some of the inherent flaws. Collaborative tagging represents a means by which museum collections can be described with a broader level of description, the end result being an increase in their accessibility. Simultaneously, users of collaborative tagging systems are given an active voice in the formulation of museum collections documentation while institutional cataloging backlogs are addressed.

This paper aims to examine the typical practices of conventional museum cataloging and the flaws contained therein, and then demonstrate how collaborative tagging can address and rectify these flaws. In doing so, the origins and popular usage of collaborative tagging as well as current museum tagging projects are discussed.

TABLE OF CONTENTS

1. Introduction	4
2. “Out of Hand” – Persistent Issues in Museum Cataloging	6
3. Collaborative Tagging – A New Democracy	17
4. Anatomy of the Tagging System	21
5. The Social Nature of Tagging	25
6. Tagging in the Museum	29
6.1. Steve.Museum	33
6.2. The Powerhouse Museum (OPAC 2.0)	38
6.3. Blind vs. Viewable	42
6.4. Sets vs. Stand-Alone	44
6.5. Relevance of Terms	45
6.6. Making the User-Resource Connection	47
7. Conclusions	50

1. INTRODUCTION

Since the turn of the 21st century museums have undergone an accelerated transition from spectators to active participants in the Digital Age. The Internet, once a space for read-only communication, has in recent years exploded into a fully participatory forum. This change, often referred to by the broad heading of “Web2.0,” is characterized most strikingly by a blurring of the once-rigid lines that separated Internet users and content providers.¹ The exchange of words and ideas that is now commonplace on the Internet is a truly democratic innovation in which restrictions on participation are, for better or for worse, almost nonexistent. Despite a steady increase in the extent to which museums have incorporated the Internet into their programming and other daily activities, a certain level of trepidation has accompanied the response of museums to the Web2.0 shift. The dual concepts of user interaction and democratic design that characterize Web2.0 strike at the very heart of the traditional museum-visitor paradigm, a model that many museums remain hesitant to abandon. Despite any misgivings on the part of the institution about alterations to the standard modes of interaction, the collaborative elements of Web 2.0 offer museums the ability to transform the manner in which they carry out their missions and serve their patrons.

One of the most visible products of the practical application of new technologies by museums is the increased digitization of collections. The ability to, through the web, search and view portions or even the entirety of museum collections is now commonplace. In addition to collecting, conserving, and displaying artifacts of

¹ van Hooland, Seth. “Spectator Becomes Annotator: Possibilities Offered by User-Generated Metadata for Image Databases.” in *Proceedings of the CILIP Cataloguing & Indexing Group Annual Conference*, Norwich (United Kingdom) 2006.

significance, museums also exercise intellectual control over their collections. That is to say, museums determine and define the significance of their collections. A fundamental element of the museum's role in society is to not only preserve and display collections, but also to express important related information to patrons. While there is little question that widespread digitization has augmented the role of the museum as a center for education and research, the speed of digitization initiatives has outpaced the ability of museums to adequately provide access and maintain intellectual control over the resulting digital collections. The primary reason for this is that the description and categorization of digital collections is commonly accomplished through standard cataloging processes. These methods of cataloging are essentially legacy practices that, though useful in some ways, are unable to meet the needs of the new global audience attempting to locate and use digital collections. Therefore one of the most salient areas in which the impact of the internet on traditional museum practices would be ultimately beneficial to museums and their patrons is that of collections cataloging in the digital realm. This paper will examine the flaws inherent in the methods many museums currently employ to catalog digital collections as well as highlight the ways in which collaborative initiatives such as social tagging stand to rectify these issues.

2. “OUT OF HAND” – PERSISTENT ISSUES IN MUSEUM CATALOGING

Among the most popular collections management software suites currently in use is a product by Gallery Systems known as The Museum System. This computer program boasts a number of modules, each dedicated to the most important aspects of collections management such as provenance, loan history, etc. Significantly, the first module, and the one that must be completed before an object record can even be created is devoted to a detailed description of the object itself. Fields for precise physical data such as dimensions and medium are provided, as are subdivisions for more intellectually oriented information such as creator, classification, date, and culture. Though the template for entering this data is integrated into a collections management database environ, this first module of The Museum System is essentially a collections cataloging tool.

A second Gallery Systems product, EmbARK, is even more highly specialized in the options it provides for attaching descriptive terms to an object. This program links itself to controlled vocabularies such as the Getty Art and Architecture Thesaurus or the Library of Congress Name Authority File, which aim to create a standard terminology to facilitate the sharing of information between cultural institutions. Though the media involved has shifted from pen, paper and index card to computer files, cataloging clearly remains a priority and a necessity for museums. Despite advancements in nearly all aspects of museum operations, the approach museums take to cataloging has remained relatively stagnant.

In stark contrast to the familiar windows-based applications offered by Gallery Systems is the Metropolitan Museum of Art’s central catalog, now administered by the Watson Research Library at that institution. The central catalog, painstakingly written by

hand on thousands of index cards in the first half of the 20th Century and now available only to researchers by request is the historical antecedent to The Museum System. In terms of the object information the central catalog contains, the results are nearly identical to those produced by modern computerized methods. Perhaps even more significantly, the process of creating a catalog record with The Museum System is nearly identical as it was for a cataloger at the Met working a century ago. In both cases a staff member, be it a curator, registrar, or assistant, must sit with the object close at hand, study it, and based upon visual cues and personal knowledge assign descriptive terms and basic information.

This type of “object in hand” cataloging employed at museums has its roots in the library field, which has for nearly two centuries employed such methods to describe bibliographic works with few changes or alterations to the practice. The shift from handwritten card catalogs to computer programs has been mirrored in that field, with most library science students cutting their teeth on cataloging programs such as Ex Libris’ Voyager system. Archives as well use these types of programs to catalog collections and records groups, and have recently taken steps towards developing fully integrated collections management systems akin to those found in museums.²

Across these separate but similar fields, there are three persistent issues inherent to the “object in hand” method of cataloging. Though steps have been taken across the board to resolve all of these problems, they remain particularly thorny for museums for a variety of reasons relating to the nature of museum collections themselves. The first of

² One such program currently in the testing stages, The Archivist’s Toolkit, is nearly identical to The Museum System albeit with some archives-specific alterations. Currently most archives use Voyager library system for collection-level description, and other database programs for maintaining accession records and provenance information.

these issues is standardization. Philosophically, all cataloging is based upon the singular premise that classifying and describing an object, be it a book, a set of papers, or a painting, will ultimately assist someone who is seeking that object or objects like it. The information created when an object is cataloged can be likened to a map, designed to guide a user to their final destination. Problematic however is that institutions, and even individual catalogers, are often of different mindsets when it comes to the means of conveying basic descriptive information. Units of measurement, spelling of names, and preferred formatting of subject terms are all open to wildly varying interpretations based upon personal experience and geography, among other factors. Libraries have perhaps been the most successful in dealing with this matter. As objects produced en masse with many identical copies worldwide, the cataloging of books most blatantly illustrates the need for a standard vocabulary for names, titles, and descriptive subject terms. As such, libraries were quick to develop such tools, the aforementioned controlled vocabularies. The Library of Congress authority files are the most visible outgrowth of this need for standardization. Compiled by librarians all over the United States, the authority file contains a listing of the standardized spelling and usage of proper names and subject terms. Use of the file in practice demonstrates how effective this can be when employed properly. A cataloger holding a copy of F. Scott Fitzgerald's The Great Gatsby may look to the cover of the book for the proper spelling and format of the author's name. However in earlier editions of the work, the author is listed as Francis Scott Fitzgerald. Still in other printings and works, he is listed as Francis Scott Key Fitzgerald. Complicating the matter further is the fact that there are no less than five other Francis S. Fitzgeralds who have authored books or articles in their lifetimes. By using the Library of Congress Name

Authority File (LCNAF), the cataloger would be able to locate the standardized version of the name, thereby ensuring that the cataloged work would be listed with others written by the same author.³

As it relates to museums however, the issue of standardization is considerably more complex. The development and implementation of a standard vocabulary in libraries has been successful for two reasons. The first of these is the aforementioned mass production of books as identical objects. The second is the comparative success of libraries at global networking. As compared to museums, the exchange of information and ideas by libraries across the world has moved at an accelerated rate. Government sponsored organizations such as the Library of Congress and professional organizations such as the American Library Association have, in the digital age, given rise to networks such as the Research Libraries Group and the Online Computer Library Center, organizations which endeavor to create and maintain worldwide global library catalogs.⁴ A tool such as the authority file, which almost seems impossible to be of any practical use without the connectivity of the Internet, actually predates computers themselves. Prior to computer networks, the Library of Congress Authority File was distributed to participating libraries in bound volumes, and updated via periodically issued bulletins.

Unlike published works, museum collection objects are ostensibly one-of-a-kind, and therefore resist such efforts to be cataloged identically. So-called “copy cataloging” in libraries refers to the practice of importing a bibliographic catalog record from another

³ The standard name is Fitzgerald, F. Scott (Francis Scott), 1896-1940.
[<http://authorities.loc.gov>]

⁴ Note that RLG and OCLC have recently merged, resulting in a single group with a global catalog (WorldCAT) containing the holdings of over 60,000 libraries.
[<http://www.oclc.org>]

institution, thereby creating a clone record for the copy of the published work at one's own institution. In the world of museums such a practice is impossible, and often unthinkable. Each object must be cataloged individually according to its own unique qualities. Perhaps even more problematic, there is no single accepted standard of names or terms for museums. The LCNAF contains the names of published authors and individuals who themselves have been the subjects of published works. Though it contains the names of certain artists, it is primarily not a resource for such information. Likewise, the descriptive terms found within its sister catalog, the Library of Congress Subject Headings, are designed to describe the subjects of published works, and are in most cases not relevant to museum objects, which demand classification and description based upon origin and physical qualities. Only recently has an initiative taken shape to create an analogous set of databases for museums in the form of the Getty Vocabularies. The Getty Vocabularies consists of three databases (dedicated to art terms, artists' names, and geographic locations) which aim to facilitate the description of collection objects by museums. Like the Library of Congress authority files, the Getty Vocabularies are arranged hierarchically. Broad descriptive terms are subdivided into more specific formats when such possibilities exist. For example, a search in the Getty's Art and Architecture Thesaurus for "videotapes" will also offer the more specific "VHS" and "Betacam." The terminology is not merely related to physical characteristics, but also to genre. The broad "Painting" can be further specified into "portrait," "landscape," "still life," etc.

For many years it could be argued quite successfully that the development of a standardized vocabulary for museums was a marginal concern at best. Museums have, for

much of the twentieth century, existed as isolated institutions when it came to their role as centers of research and intellectual caretakers of collections. The proceedings of organizations such as the American Association of Museums or the International Council of Museums show that, despite some concerted efforts to formulate codes of ethics and operating standards for physical care of materials, decisions on cataloging were largely left to the individual museum. As object descriptions were primarily for in-house use, their ability to be understood by outsiders was of little relevance. The raised profile of individual museums on the global stage as made possible by the Internet, plus a newfound concern for accurate and publicly available provenance information has changed all of this. Aside from the obvious benefits of increased access, there is now a distinct *need* for museums to have their collections easily searched and located. The Getty Vocabularies is but one initiative designed to aid in fulfilling this need but the overwhelming lack of standardization across the museum field and the slow speed with which museums approach retrograde cataloging projects still present obstacles. While museums have much to gain from exploring new methods of cataloging, practical concerns related to funding often prevent such exploration. Put simply, though more complete cataloging enhances the intellectual value of collections and opens the door to a specific set of museum users, it does not drive attendance or create income. At many museums cataloging projects must, for very pragmatic reasons, take a backseat.

For as many questions as a tool like the Getty Vocabularies portends to answer, it raises several others. Above all, how effective is museum cataloging in conveying the significance of museum objects? At the heart of this skepticism is the second major issue related to utilizing the “in-hand” method to describe museum collections. Bibliographic

cataloging in a descriptive sense is focused upon subject matter and content. “What is it about?” This fundamental question can be applied with considerable success to almost any printed work. It forms the touchstone of the entire bibliographic cataloging process. However the cataloging of museum objects is far more intricate, and for most museum professionals and scholars alike the “what is it about?” test is of little use for describing collections due to their varied nature. In museum cataloging, other queries take its place. “What is it made of?” “Who made it?” “When was it made?” Here again, the uniqueness of museum collections poses problems that are often farcical in nature. Robert Rauschenberg’s “Canyon” is made of oil, house paint, pencil, paper, fabric, metal, buttons, nails, cardboard, printed paper, photographs, wood, paint tubes, mirror string, a pillow, and a bald eagle. The gun that John Wilkes Booth used to assassinate Abraham Lincoln was manufactured by the Derringer Company of Philadelphia. A meteorite fragment fell to Earth thousands of years ago.⁵ As these examples show, when the typical museum cataloging questions are posed, the answers rarely illustrate significance of the objects. In each of the above examples, very little is revealed which would indicate why each object is part of a museum collection. One might wonder then why such attempts should be made at all, if the very nature of the objects themselves strikes at the usefulness of cataloging. The obvious answer is that it is not the objects themselves, which are the obstacles to a more overtly useful type of museum cataloging, but the methods used by catalogers.

Finally, there is a third concern, which itself is interlaced with those already mentioned. Within libraries, archives, and museums there exists a considerable semantic

⁵ Objects and descriptions taken from the National Gallery of Art, the Ford’s Theater Museum, and the American Museum of Natural History, respectively.

gap between those who catalog and formulate descriptions and between those who they are intended to serve. The most common complaint against the Library of Congress Subject Headings (LCSH) is that the descriptive terms it contains are often arcane, obtuse, and arbitrarily designed, which in turn hinders the effectiveness of the library catalogs that rely on them for cataloging. The immediacy of the Google search engine and the way it has transformed approaches to finding information only serves to magnify this concept. The specific nature of museum objects plus the specialized knowledge of curators, registrars, and scholars who work with them combine to underscore this issue in the world of museum cataloging. Even if a standardized vocabulary were implemented for museum cataloging, it could in the end hinder, rather than enhance access.⁶ Improper use of terminology, unclear categories of information, and misspellings of names all pose obstacles to the user in these instances.⁷ To hearken back to the example of F. Scott Fitzgerald, how could an art museum cataloger hope to avoid confusion when cataloging the works of medieval or Byzantine artists whose names are sometimes recorded as only a single surname, and whose biographies are nonexistent?

These issues are not new, and have in essence existed since the first museums began to catalog their collections centuries ago. However, they are brought to the forefront and further complicated by the fact that the digitization of museum collections has become commonplace in the modern age. Digitization, in the most basic sense, simply implies the transformation of information into a new format. For museums that

⁶ Harpring, Patricia. "How Forcible are Right Words! *: Overview of Applications and Interfaces Incorporating the Getty Vocabularies." in *Proceedings of Museums and the Web Conference: 1999*.

⁷ Trant, Jennifer. "Understanding Searches of an On-Line Contemporary Art Museum Catalog." *Archives and Museum Informatics*. Fall 2006.

are digitizing their collections (and ostensibly, the information related to them) the end result is typically the dissemination of such information to remote users via the World Wide Web. The technical processes by which museums can digitize their collections are numerous, however the final product of digitizing collections is most commonly a remotely accessible database that contains object records and visual representations of them. It is in this context that the true necessity of revisions to the museum cataloging process is revealed.

At first glance the benefits to the digitization of museum collections are numerous. The ability to serve a wider pool of patrons across the globe is perhaps the most obvious result, but increased interest in the museum and newfound awareness of the collections on the part of patrons and staff alike are also commonly cited.⁸ For museums, the digitization of collections enables institutions to carry out all of their most basic activities that are contingent upon the data-supported, visual representation of collection objects in a global theater. This includes exhibitions, programming, and research. The benefits of digitization do not come free of any attachments, however. When a museum object is digitized in the form of a digital image or database file, it becomes for all intents and purposes an abstract set of data, isolated of any relationship to the original physical object whatsoever. The only way to convey this relationship with any effectiveness is through metadata; data that classifies, categorizes, and otherwise identifies the digital object and conveys significant information about it. Given the high costs and staff involvement required to undertake full-scale digitization projects however, the creation

⁸ Fuller discussion of the benefits of digitization projects can be found in the “Status of Technology and Digitization In the Nation’s Museums and Libraries 2002 Report” compiled by the Institute of Library and Museum Services. [http://www.ims.gov/publications/TechDig02/]

and application of the necessary metadata is all too often treated as an afterthought. In almost all cases metadata creation is approached from the traditional museum cataloging perspective.

This tendency to integrate traditional cataloging methods can be seen in the literature regarding many digitization projects. Within such initiatives, the primary concern is often the technical format of the metadata. Efforts generally focus upon negotiating a shared encoding standard that can be understood across museums. Technical matters, such as creating data in a form that can be translated into new formats such as XML or Dublin Core frequently tends to overshadow the data itself, the source of which is almost always the in-house collections management system.⁹ However as was noted, programs such as The Museum System originated as systems for staff use. They are not centered upon the user outside of the museum, nor were they ever truly intended to be. In reference to the pre-existing issues already mentioned the product of this “digitize first and ask questions later” approach is a confusing set of digital objects that are ultimately less useful than they would first appear. Though the potential for access is greatly heightened, the provided metadata acts as the only possible means of locating and identifying the collection objects one is seeking. When the terms of description encoded as metadata are singularly formulated according to specifications, which themselves are less-than-ideal for museum collections, attempting to access digital collections can be compared to crossing a bridge while blindfolded. Reaching the other side will be the result of a haphazard process of trial and error, if it ever occurs at all.

⁹ Elings, Mary W. and Gunter Waibel. “Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives, and Museums.” *First Monday*. Volume 12, number 3 (March 2007)

In light of this dilemma, museums must seek to develop new ways of cataloging digital collections. The resulting methods must, above all, balance the needs of the institution with those of the user and demonstrate a shift in priorities from the cataloging process to the final product. Social tagging, a Web2.0 practice borrowed from decidedly non-academic institutions, offers some answers.

3. COLLABORATIVE TAGGING: A NEW DEMOCRACY

Collaborative tagging, alternately known as social tagging, social bookmarking, folksonomy, or social classification refers to a process by which individual users contribute subject terms to digital resources based upon their own singular conceptions of the resource's values. The resulting descriptive information is then compiled and associated with the resource, usually with respect to patterns of consensus and repetition that emerge in the accumulated data. Collaborative tagging is an intriguing development in part because of the processes that go into it, and in part because of the information that is its final product. The descriptive vocabulary that is the inevitable end result of collaborative tagging is the collected outcome of many minds, perspectives, and thesauri. Contributors and users are drawn from a single pool, and ostensibly the resulting terms of description will be of greater use to wider audiences.

To best illustrate how collaborative tagging works in practice, a direct comparison with subject indexing is useful. In typical subject indexing, the method of encoding terms of description is heavily controlled by an encoding standard. Additionally, the terms themselves are chosen from a controlled hierarchical vocabulary. A controlled vocabulary is a subject-indexing tool in which the descriptive terms are chosen from a predefined set, determined by the vocabulary's designers. Most controlled vocabularies also have some type of hierarchical component, in which terms are subdivided to allow for a greater level of specificity in description. For example, a museum cataloger employing the hierarchical vocabulary of the Art and Architecture Thesaurus (AAT) might mark a nineteenth-century painting as such: "Painting – 19th Century." A subject term search for "painting" would find the work within a set of results containing any

number of nineteenth-century paintings. Since the chronological descriptor is a subdivision applied to the main term of “painting,” a search for the nearly identical “nineteenth-century painting” would not return the same results. Conversely, in a typical tagging system, “painting” and “nineteenth century” would exist as separate tags, and therefore searches for “19th century painting,” “painting 19th century,” “painting,” and “19th century” would all be useful to some degree for locating the work in a digital repository. Additional terms such as “oil painting” or even more object-specific adjectives describing the work’s physical appearance or historical context (“Folies Bergere,” “bar,” “impressionism,” “Manet”) could also exist as tags. Since the vocabulary is uncontrolled, the possibilities for tags are almost limitless. The greater the number of tags that are applied to the digital resource, the more accessible it inevitably becomes. This rise in accessibility is multi-directional, as it targets not only users who are seeking one specific object, but also those who may simply be browsing entire collections.

Significant discussion of the potential advantages and drawbacks of collaborative tagging for museums warrants a look at the process itself, as well as its prevalent uses. The basic acts involved in tagging are quite simple, and are not artificial constructs of the digital age. Forming the touchstone of tagging systems is an innate human tendency to classify and describe that manifests itself in many ways, of which tagging is only one approach. Devoid of any discussion of its greater psychological or social implications however, what is “tagging?” To “tag” simply means to apply a single-word descriptive identifier to a resource or object. For the purposes of this discussion, the resource is almost always understood to be digital. In tagging, the source of the descriptive term is

uncontrolled, and the tags are limited only by the individual tagger's conceptions, accumulated knowledge, and vocabulary. The tagger is generally not required to indicate in any way how a tag relates to a resource. As such, tags can refer to a resource's subject, physical qualities, point of origin, or other attributes. Once collected, tags are then made searchable. In most tagging systems there is no limit to the number of tags that can be applied to a single resource, and ostensibly the greater the number of tags applied to a resource the more accessible that resource will become. While the same could be said for traditional subject indexing, the lack of control over the ontology and encoding of the tags is what separates tagging from traditional cataloging both in practice and in purpose. Such free reign on the part of the user and the accompanying emphasis upon personal knowledge and insight also fosters elements of community within collaborative tagging systems.

The types of websites which employ collaborative tagging demonstrate the myriad of realms in which the practice can be applied, and also give some hint as to its potential for museums. Del.icio.us [<http://del.icio.us>] allows users to tag their own personal collection of bookmarked web pages, which are then made available to other users of the tagging system. Photographers, both amateurs and professionals, use Flickr [<http://www.flickr.com>] to upload digital photographs, which are then tagged and made available to other users. Internet video clearinghouse YouTube [<http://www.youtube.com>] allows its users to tag videos, which they upload. In each of these cases the resources being tagged are inherently digital; however this need not be the case. Last.fm [<http://www.last.fm>] has its users tag their favorite songs and musical artists, resources which exist outside of the realm of the Internet. Similarly, LibraryThing

[<http://www.librarything.com>] and GoodReads [<http://www.goodreads.com>] have users tag the titles and authors of the volumes that comprise their own personal libraries.

Though seemingly disparate, these differing types of resources (photographs, songs, books, videos, blogs, news articles, to name but a few) are united by a single ideological thread: their significance varies from person to person, from community to community. The aggregated results of collaborative tagging reveal insights into the very nature of the resources they describe. In many cases these insights are highly specific, and might otherwise go overlooked if the task of description was left solely to professional catalogers. It is for this reason that collaborative tagging is a viable means of collecting, assembling, and disseminating metadata, and stands to become one of the most powerful classification tools in a new digital age. As barriers to participation fall in the face of Web2.0, the number of digital resources available on the web multiplies exponentially. Ensuring that these resources are not “lost in the crowd” demands a new approach to classification and description.

4. ANATOMY OF THE TAGGING SYSTEM

In spite of their increasing prevalence, the dynamics of tagging systems are not yet widely understood. While the processes involved make them relatively simple to implement from a technical perspective, formulating a model of behavior within a tagging system and using it to predict eventual results is far more challenging. Marlow identifies three distinct elements of a tagging system: resources, users, and tags.¹⁰ Others identify the system elements as objects, humans, and terms, somewhat less ambiguous terminology.¹¹ Nomenclature disputes aside, an exploration into the elements that comprise the tagging system and how the system itself factors into the act of tag creation is worthwhile.

Within the context of collaborative tagging systems, the resource (object) is that which is the target of description, and to which the tags will be associated. The resource can be supplied by the system itself, or more commonly, it is supplied by the users. “In reality, any object that can be virtually represented can be tagged or used in a tagging system.”¹² This notion of virtual representation is an important element in the definition of a resource. Digital objects have a dual existence, as they are both resources unto themselves, and also representations of an object. A digital photograph is, in its most basic sense, a computer file composed of encoded 1s and 0s. Beyond this however, it is more widely understood to be a virtual representation of an image. The ability of a web browser to transform the file into something that is visually recognizable directly links

¹⁰ Marlow, C., Namaan, M., Boyd, D., and Davis, M., “HT06, Tagging Paper, Taxonomy, Academic Article, ToRead”, Proceedings of the 17th Conference on Hypertext and Hypermedia: 2006.

¹¹ Mesnagi, Cedric, and Jazzari, Mehdi. “Specifying the Collaborative Tagging System.” Proceedings of the Semantic Authoring and Annotation Workshop: 2006

¹² Marlow, 2006.

the file to what it represents. In some cases however, such linkage between the resource and what it represents is more abstract and requires a greater leap of faith. In the LastFM system it is understood by users that that song titles and artist names serve as virtual representations of songs and artists. Thus, when tags are applied to the names, they are in essence being used to describe actual songs and artists. Similarly, on LibraryThing tags are applied to book titles, which represent printed books. This common knowledge is implicit across the entire system. All tagging systems hinge upon an understanding of this implied link between the virtual and the real. Regardless of the manner in which it is conveyed, recognizing this link is a responsibility that falls upon the shoulders of the user.

In the tagging system the user is always a human being. However just as each tagging system centers around one chosen type of resource, the design of tagging systems can also play some part in cultivating users. There are essentially two approaches to user control employed by tagging systems. The first, and more restrictive, is *self-tagging* in which users are only permitted to tag resources that they have created. The second approach is *free-for-all tagging*, in which any user may tag any resource. Certain tagging systems, most notably Flickr, have experimented with a third approach, sometimes referred to as *permission-based tagging*, in which the user responsible for the resource creation also has control over tagging rights.¹³

The final element of the tagging system is the tag itself. The only uniform characteristic of the tag itself is that it is uncontrolled, and not drawn from a standardized vocabulary. Despite this relative freedom, the tagging system can affect tag generation

¹³ Marlow, 2006

and aggregation by altering the arenas in which the act of tagging takes place. *Blind tagging* refers to tagging systems in which users are not shown other tags that have been attached to a resource. *Suggestive tagging* is a system in which likely tags (determined by aggregating the efforts of previous taggers) are suggested to the user at the time of tagging, and *viewable tagging* refers to a system in which all previously applied tags are shown to the user. The implications for vocabulary generation are strong in each instance, and pending the desired results, a case can be made for the relative benefits and drawbacks of each scenario.

It is evident that the relationship between the user, the resource, and the tag is the driving force behind collaborative tagging. It is also clear that the way in which the system relates to each of these elements will in some way have an affect on the eventual production of data. However, what other factors integrated into the tagging system's design come into play during tag creation? One of the most salient of these is the tagging system's approach to tag aggregation, that is, how it collects data from the tags that are supplied. One approach to this, the *bag-model*, allows for a multiplicity of tags for any given resource. In terms of data collection, this means that a repeated tag on a resource will be collected multiple times. In practice, this means that certain repeated tags can be identified as "popular" relative to others. With many users participating, the information gathered from the *bag-model* can give some clue as to feelings of consensus among users, and of user behavior and community development. Conversely, the *set-model* does not allow for such multiplicity. In *set-model* aggregation once a tag has been harvested, it will no longer be recognized. While this limits the system's ability to assess the popularity of certain tags among users, with the opportunity for repetition removed it can

also spur the creation of new tags, thereby broadening the tag pool.¹⁴ Any design of a collaborative tagging system must consciously make decisions on such matters with regard to the intended purpose of the system.

A final factor to consider within the collaborative tagging system is resource connectivity. Resource connectivity refers to the way in which separate resources relate to each other devoid of tags. In some systems, resources deemed similar (by the system, or in some cases the resource creator) are grouped together. These groupings, or the lack thereof, can ultimately have an effect on tag creation. As most museums tend to group collections based upon one or more shared cultural, artistic, or thematic contexts, the concept of resource connectivity is one which inevitably plays a role.

¹⁴ Marlow, 2006

5. THE SOCIAL NATURE OF TAGGING

Discussion of solely the technical processes of tagging fails to clearly identify the underlying social aspects of the process. It is the collaborative and inherently social nature of tagging, which makes it such an exciting development and which most strikingly separates it from traditional cataloging. Mesnage and Jazayeri perhaps best express the value of collaboration:

The collaborative tagging system allows anyone to index any object without any restriction with meaningful terms. The notion of meaningfulness is subjective, but the combination of points of views from many people is more complete than the one of only a chosen subset (experts).¹⁵

The collaborative aspect of tagging is therefore a natural outgrowth of user participation across a broad spectrum. To fully grasp the nature of this outgrowth requires an understanding of why users participate in collaborative tagging systems in the first place. There are several competing schools of thought on this topic, none of which fully explain the popularity of tagging, but which nonetheless provide some insight. The explanations range from the abstract and the ideological to the specific and practical. Perhaps the broadest of these explanations is the notion of the reversal of authoritarian control over information. As mentioned previously, collaborative tagging implies a certain reordering of the information distribution paradigm. In the context of the collaborative tagging system readers become authors and tags entered by experts have as much value as those entered by amateurs. Subjectivity is wholly embraced and in many ways the ideas of “correct” and “incorrect” are negated. Writes David Weinberger, “An author is an authority when it comes to what she intended her work to be about, but not

¹⁵ Mesnagi, 2006

about what it means to others. When it comes to searching, what a work means to the searcher is far more important than the author's intentions." The level of *trust* bestowed upon the participant in this type of system is nearly unheard of, particularly in such an academic realm as a museum or a library. As Weinberger artfully expresses, "It [tagging] sticks it to The Man, especially if The Man happens to be a traditional taxonomist."¹⁶

Weinberger's colloquialism points to a broader concept, the overtly democratic nature of collaborative tagging. The collaborative tagging system (most obviously in the *free-for-all* model) is one in which representation is not limited to a single group of users, and is not based upon expertise in the resource area or any other learned aptitude. Rather, participation is guaranteed. Mesnage and Jazayeri assert, "As the number of people working collectively grows, the system of interaction between them must be simplified. In a democratic political system, the voting system is an example of a simple interaction to decide collectively at a large scale. To organize knowledge together, the interaction system between humans must be as simple as possible." The overall simplicity of the tagging process and its ability to offer a collective voice is understandably attractive to many users.

Marlow also identifies several somewhat more specific user incentives that motivate users to participate in collaborative tagging systems. They include opportunities for future retrieval of resources, the contribution and sharing of personal knowledge or insight, and the tagging system as an arena for expression of opinion, self-presentation, and the attraction of attention. With the exception of resource retrieval, nearly all of these incentives are heavily informed by the notion of the user pool existing and relating as a

¹⁶ Weinberger, David. "Tagging and Why it Matters." 2005.

community. Some, such as attracting attention and self-presentation, can exist only within that context.¹⁷

This idea of community represents a third motivator for user participation. Collaborative tagging systems possess the ability to represent existing communities, and perhaps more impressively, to give rise to new ones. Flickr provides an illustration of both simultaneously. Among Flickr's more intriguing developments is the "Geotagging" feature. Geotagging refers to a process by which users add a geographic tag to their photos, indicating where the photograph was taken. Suggested tags are country names, city names, or regions. Through Geotagging, pre-existing local communities are instantly and seamlessly integrated into the collaborative tagging system. Users can view photographs taken in a single place, and connect to users in a specific region. Appropriately however, the Geotagging tool now integrated into Flickr is one that was developed independently and suggested by a group of Flickr users from all over the world who shared a belief that such a practice would be useful and interesting. The Flickr Geotagging community is therefore one that arose organically in the system. A look at the other groups that exist on Flickr reveals a mix of such organically formed groups and pre-existing communities that have co-opted the tagging system. The "Canon EOS 5D Group" is for users worldwide who use the same model of digital camera. The "JPG Magazine Group" is for subscribers of a certain print magazine. The "John Mayer" group is for fans of a specific musical artist, while "Lighthouse Lovers" is oriented to those with an affinity for seaside beacons. Community formation in the tagging system is limited only by the distinct personalities and desires of the users, and once communities have

¹⁷ Marlow, 2006

been formed, the exchange of tags and resources both between and within these communities takes on an entirely new relevance.

6. COLLABORATIVE TAGGING IN THE MUSEUM

In light of the wide array of resources collaborative tagging can be used to describe, it is not difficult to imagine the application of such a system to digital museum collections. In museums, many of the potential benefits attached to collaborative tagging in other arenas are magnified. The increased access to digital museum collections that collaborative tagging can provide would only serve to supplement their existing educational and research value. There are however, several caveats museums must bear in mind before wholly embracing collaborative tagging. As Trant points out, “Tagging in a museum context may differ from other implementations of social tagging (including the shared bookmarking services such as del.icio.us and Connotea) because museums have existing relationships with visitors that define a social and cultural context for tagging activity.”¹⁸ When discussing tagging and museums, two of the key concepts of the collaborative tagging system bear closer scrutiny: the resource and the user.

Within a museum collaborative tagging system the resources take the form of digitized museum collections. The act of describing any object from a museum collection, through cataloging or other means, is heavily dependent upon visual recognition. Therefore, if users are to carry out tagging as related to museum collections the accompanying resources must come in the form of a digital image. While subjectivity among users is a fundamental component of the tagging system, collaborative tagging’s legitimacy as a means of description is dependent upon the tagger having an accurate conception of the resource. The essence of a Warhol, a Kandinsky, or a Seurat simply

¹⁸ Trant, Jennifer. “Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept.” Draft for *New Review of Hypermedia and Multimedia*: 2006 May 31

cannot be conveyed through a textual description. This need for digital images raises a number of practical concerns. Digitizing museum collections in a manner that is efficient is a costly enterprise for museums. Even institutions that have made great strides in the area of digitization have commonly fallen short of digitizing *everything*. The issues that hinder digitization run the gamut from the technical (What file format? What resolution?), to the legal (Who owns the image? How can it be used?). The burden for all of these decisions falls on the museum. This is unlike in a system such as Flickr or YouTube, where resource creation is almost entirely undertaken by the user community, or like Last.FM, where the practical concerns of digitization are bypassed completely.

Apart from the challenges of initially setting up such a system, there are other issues museums must contest with even after resources have been created and are made available for tagging. The inherent educational value of museum collections lay in their historical and social context. Museum exhibits, documentation, and educational programs focus on far more than the mere aesthetics of collection objects. It stands to reason that collaborative taggers should follow suit. If collaborative tagging in a museum is to be a productive and worthwhile endeavor, the nature of museum objects and the educational missions of museums demand that a virtual representation of a museum object somehow convey more than a simple visual cue. This is where the issue of user identity comes to foreground. While certainly some users will be familiar with the background and context of highly recognizable works of art, most museum collection objects are immediately well-known to only a small segment of the population. This is the same segment that has traditionally controlled cataloging in the past. The museum must therefore somehow attract new users and promote useful and substantive tagging without at the same time

inserting the curatorial voice into tag formation. The uncertainties surrounding this question lay at the heart of many of the theoretical doubts regarding collaborative tagging.

At first glance, few of the specified user incentives translate effectively into a museum tagging environment. The notion of tagging museum objects (versus, for example, a song) as an expression of personal opinion or to attract attention seems antithetical to the goals, which museums might set forth when participating in a collaborative tagging project. These goals would likely include collecting and dispersing collections information, finding new ways of thinking about objects and their educational value, and increasing accessibility to the collections by tapping into new communities of patrons. Only the user incentive of tagging as a method of sharing personal knowledge dovetails with this notion of tagging as facilitating collections-based education. But without user incentives that have a broad appeal, attracting participants to the collaborative tagging system from outside of the existing museum and art history communities is a challenge. Though the voices of these communities are important in many ways, they are populated by the same scholars and professionals who are currently responsible for most museum cataloging. It is “outsiders” who have the most to gain from new methods of description, and whose participation is vital to the success of tagging.

Notwithstanding these looming issues, a number of museums have initiated collaborative tagging projects. These projects range from small ad-hoc efforts to large consortial ventures, and each project has made its own choices on features such as layout, presentation, and user options. Unfortunately, most of these initiatives are still at such an early stage that it is difficult to track data that would point to the long-term benefits of

collaborative tagging in the museum context. Despite this the project interfaces themselves, when considered along with the minimal data available, can be a valuable source of information. By comparing the methods and findings of two of the most well developed and monitored collaborative tagging projects there is much that can be gleaned about how the design of the tagging system itself ultimately affects tag formulation and description. Once this notion is more fully understood, the value of collaborative tagging in relation to traditional cataloging for museums can begin to be assessed.

6.1. STEVE.MUSEUM

By far the largest ongoing attempt at collaborative tagging in museums is the Steve.Museum project. Designed in 2005 and first made available to the public in 2006, Steve is a cooperative enterprise of several major museums and other cultural organizations, including some larger names such as the Metropolitan Museum of Art, the Guggenheim Museum, and the Los Angeles County Museum of Art.¹⁹ Before examining the interface, it is worth noting that Steve is a grant-funded project, the ultimate underlying purpose of which is to collect data that answers many of the questions that have been posed about collaborative tagging. This directness of purpose is in some ways a double-edged sword. The Steve tagger is an exercise, manipulated and overseen with an eye for data collection. The differences between this type of controlled project, and a collaborative tagging project taking place in a museum context (versus a research context) are subtle but important.

At the heart of the Steve project is an understanding of the semantic gap between art historians and the average museum patron, a notion explicitly set forth in the project documentation.

Together, working within steve, we hope to learn more about how to improve access to museum collections. We've realized that what people remember about a work of art often isn't described in museum collections documentation, and are aware that the vocabulary of museum professionals (curators and registrars) isn't always that of the general public. We want to explore user-contributed terminology, collected on the Web, as a way to address these gaps. User-created descriptions could provide museums with missing subject-based information for their collections databases and make our on-line resources more useful.²⁰

¹⁹ A full list of Steve.Museum participants is available on Steve's FAQ page at: http://steve.museum/index.php?option=com_content&task=category§ionid=20&id=38&Itemid=66

²⁰ Steve.Museum FAQ

Interestingly, the Steve project recognizes that “User-based description also allows people to develop a more personal relationship to museums and museum collections.” Therefore, from the outset, both the utilitarian goal of increasing access to collections and some interesting secondary incentives related to personal identity and potentially also community development are suggested. This fostering of personal relationships is something of a composite incentive, most similar to that of “future retrieval” but also linked to the ideas of opinion expression and self-presentation. While the idea of a museum undertaking a collaborative tagging project to augment the formulation of user identities may seem contrary to the museum’s goals, to the individual user, living in the age of Myspace, Facebook, Second Life, and other identity-based Web2.0 applications, it seems almost second nature. Ultimately, to attain the widest and most diverse possible pool of users, museums may have to make concessions when it comes to the matter of incentives and understand that the needs of the institution may not be congruent with the desires of the user.

In examining the Steve Tagger’s functionality, it is vital to understand the aims and understandings of the designers, and how these translate into the system’s design. As a shared project between several different institutions, the Steve Tagger’s interface contains several elements that are somewhat unique among collaborative tagging tools.²¹ In the current version of the tool, users wishing to tag artworks are presented with an introduction screen upon reaching the site. Here they are given the option to register, or if they wish, to simply bypass the registration process and tag artworks. Once a selection

²¹ The Steve Tagger is a piece of open-source software and therefore is constantly evolving. The version tested and discussed here is Release 2.0, dating from October 18, 2007.

has been made and (if necessary) registration completed, the user is presented with a group of seemingly randomly generated terms and asked to select more than one term before proceeding. Though it takes some time to recognize the purpose of selecting these terms, it is a fundamental step in the process. Each of the introductory terms is a previously supplied tag. Those that are selected determine the set of artworks that are supplied to the user for tagging. For instance, if a user selects the terms “triptych” and “pulp fiction,” the resulting set of six artworks will all have been previously tagged with one or both of these terms. This manner of assembling sets of objects for the user is not quite like any other tagging system, though it has several interesting consequences. First, as there is no clue given as to what significance the terms have, users ostensibly choose those that are the most interesting to them objectively. This leads to the tagging of objects that relate in some way to a pre-existing interest. If one was to write a dissertation about Albrecht Dürer, one might end up choosing “woodcut” from the list. Outside of the domain of art and art history, one might choose “baseball” or “sailing” or “sharks.” This is significant because of the way in which tags are typically applied. One who selects sailing and sharks may be confronted with Winslow Homer’s *The Gulf Stream*, a painting which arguably has very little to do with either topic. It is in this manner that users are connected with works which they might never otherwise find.

Perhaps the most significant feature of Steve is its cognizance of the relationship between the resource, the system, and the user. To this end, multiple environments have been integrated into the tagging tool over its history, each being made available for a set amount of time and designed to collect a specific data set. There are two main variables that change between environments, the grouping of works into sets, and the level of

resource-related information supplied to the user. In one environment, the entire set of works to be tagged is viewable together, by means of a sliding menu bar at the bottom. Users can jump back and forth between works in the set, and presumably will focus their tagging on those works that are the most visually stimulating. In the currently available environment only the work that is to be tagged is viewable at any one time, forcing a user to focus on the individual work before choosing to tag it or skip it. In another environment, the traditional museum documentation regarding the resource (artist, date, dimensions, etc) is displayed under the image. In a second environment this data has been replaced by a list of the tags supplied by other users. A third environment lacks any of this information (displaying the image only), and a fourth environment contains both elements and the image. Similarly, in one environment sets of works are assembled according to the institution where they are held. In other environments, sets are grouped by medium, or as with the current environment, by common tag. The various iterations that can be presented to the user in such an experimental and multi-faceted tagging environment are vast in number.

These manipulations on the part of Steve are useful because for each hypothesis that can be proposed regarding the relationship between tagging and the interface, there is a related counterargument. For example, regarding blind tagging versus viewable tagging it has been asked, “Does user behaviour change when they see the tags that others assign? We can hypothesize two possibilities: that users mimic what is presented to them, or that they strive to be different.”²² Or regarding resource groupings, do users view tagging as a contribution to a single institution when resources are arranged by museum and tag more,

²² Trant, et al, 2007

or does the diversity of the resources spark more valuable tagging when works are arranged by medium or at random? Though it is only preliminary, the data from Steve is illuminating. This owes much to the level of consistency provided by using the same set of resources across the different environments. Before tackling the ramifications of such data however, we will turn to a tagging project from outside of the Steve domain for a comparison of system design.

6.2. THE POWERHOUSE MUSEUM (OPAC 2.0)

With a diverse collection of approximately 385,000 objects, the Powerhouse Museum located in Sydney is Australia's largest and most popular museum.²³ Its education-centered mission touches upon science, history, and design, and its holdings include artifacts, artworks, clothing, fossils, and countless other object types. In the summer of 2006, the Powerhouse Museum launched the beta version of OPAC 2.0 (online public access catalog), a new searchable collections database that incorporated collaborative tagging as one of its central features. Here too, the semantic gap is identified, explained, and commented upon by the accompanying literature at an early point. "Sometimes museums describe objects in language that is highly specialist and user added keywords are useful in bridging the 'semantic gap' between the language of the museum and that of the user."²⁴ In relation to the notion of collaborative tagging as a democratic development, the underlying ramifications of this statement and the similar one from Steve's documentation should be noted. Each of these projects has chosen the theme of "bridging the semantic gap" as its rallying cry to taggers. But we know that this is only one of many reasons museums might undertake collaborative tagging, and in fact it is clear that the potential benefits of collaborative tagging are at least equally as great for the institution as they are for the user. Nonetheless, by publicly acknowledging the failures of traditional cataloging and admitting the inherent flaws of the one-way didactic paradigm, the user is encouraged to participate in tagging activity. Therefore, one can already identify a potential solution (albeit an overly simple one) to the question of how

²³ <http://www.powerhousemuseum.com/about/index.asp>

²⁴ <http://www.powerhousemuseum.com/collection/database/browsekeywords.php>

to encourage tagging on the part of previously excluded communities: through openness, honesty, and an appeal to the sensibilities of the user.

Before even delving into the functionality of the interface, one should recognize the numerous fundamental differences in scope, practice, and purpose between Steve.Museum and OPAC 2.0. First, the set of resources that comprise OPAC 2.0 are both narrower and more diverse than those of Steve. The digitized objects that comprise the collections of the Powerhouse Museum contain a relatively small number of artworks, and are entirely of the same source institution. This is in direct opposition to those resources that are in the collections of Steve, which are exclusively art objects, but are drawn from varying institutions. Also, the OPAC 2.0 project's goals are different from those of Steve. As mentioned, Steve has been formulated from the ground up as a research project, and has been tested, manipulated, and overseen throughout its lifetime with the purpose of testing hypotheses and drawing conclusions. OPAC 2.0, on the other hand, was implemented as a means of replacing a previous online catalogue, the records of which had not been updated since 2001.²⁵ Therefore, while the project is not so strongly focused on developing an understanding of tagging processes, it may have more to say about tagging as it relates to traditional cataloging and resource accessibility.

The interface of OPAC 2.0 is quite simple and is integrated within the Powerhouse Museum's main online collections database. Unlike in the Steve tagger, tagging in OPAC2.0 is not presented as an end unto itself. Rather, the introductory environment allows the user to search across the collections database, browse categories

²⁵ Chan, S., Tagging and Searching – Serendipity and museum collection databases. In J. Trant and D. Bearman (eds). *Museums and the Web 2007: Proceedings*. Toronto.

(supplied by the museum), or browse user tags. Terms entered in the main search box will be searched for across both user tags and categories. When choosing “Browse Tags,” the user is presented with a tool that is common in other collaborative tagging domains, the tag cloud. The tag cloud consists of a randomly generated list of terms that have been used as tags. What gives the interface its “cloud” effect is that terms which have been used as tags in multiple instances are represented in a larger/bolder font, conveying their commonality in an easy to understand visual manner and enticing further exploration. Note that there is no way to browse through a list of every user tag at once, nor is there any inclusion of a tag cluster. A tag cluster, essentially a cousin of the tag cloud, is a second visual representation in which tags are not only associated with specific resources, but with each other. If two tags commonly appear together (such as “water” and “ocean”) these will be listed together in the cluster. The tag cluster, though not present in any of the current examples of a museum collaborative tagging, is a hallmark of many other tagging systems and represents one of the most dynamic and functional approaches to the development of a new vocabulary.

If a user chooses a term from the OPAC 2.0 browse page or searches for a term from the main search, all objects tagged with that term are returned in the form of a brief list with accompanying thumbnail images. For example searching for or clicking on “Stoneware,” a relatively popular tag according to the tag cloud, returns records for 18 different objects, all of which have been tagged with the term. Once a user has selected an individual object record from the list by clicking on it, tagging may then be accomplished by means of a small and unobtrusive text box located on the side of the page.

The OPAC 2.0 interface supports an extreme version of visible tagging. Not only are previously supplied tags able to be seen when tagging, but so is any available museum-supplied information. This can be as simple as basic object data, or it can include extensive description and discussion of the object's significance. Also displayed alongside the tagging module is a list of so-called "related objects" taken from the museum's formal categories, and "related searches" gleaned from these objects' tags.

6.3. BLIND VS. VIEWABLE

One of the most pressing questions regarding tagging system design and tag generation is that of metadata presentation. The question of whether museum-supplied metadata compels users to tag more or less frequently is answered at least in some part by the findings of the OPAC 2.0 project.

The most tagged object, a badge commemorating Lambton Bowling Club's 75th Anniversary (registration number 89/591), has been tagged 10 times with a mix of aesthetic (kangaroo, marsupial, parrot, rosella), material (enamel), descriptive (membership badges, bowling club, anniversary, badge), and social (Australiana) terms. The formal object record for this badge has two images but no statement of significance, no production or history notes, and only a very basic one-line description and dimensions. In this case, the user keywords offer alternative means to discover the object given the lack of other object data.²⁶

Other frequently tagged objects from the Powerhouse Museum also lack supporting metadata. The findings of the Steve prototype tell a similar story. In that scenario, both resources consisting of only an image and resources consisting of an image with a caption averaged between 7 and 8 tags per work.²⁷ The Steve data also indicates that there was a high level of consistency among the actual tags that were used regardless of the presentation.

Perhaps even more important for this discussion, OPAC 2.0 allows users to tag digitized museum objects *without* digital images. This is something that takes place consistently throughout the system, and while a brief sampling of objects without images shows that the number of tags is generally less than those for objects that do have images, the phenomenon is noteworthy because it flies in the face of the generally accepted wisdom regarding tagging in museums. Perhaps if digital images are not as wholly

²⁶ Chan, 2007

²⁷ Trant, Proof of Concept, 2006

necessary as previously believed, one of the most cited obstacles to collaborative tagging is broken down.

6.4. SETS vs. STAND-ALONE

A second question of system design comes in the form of the presentation of resources, whether in isolation or as belonging to a larger set defined by some shared characteristic. Given the data available from both Steve and OPAC2.0, this relationship is somewhat more difficult to identify. Steve reports that in earlier versions of the tagger, when works were presented as a set, the number of tags proportionally associated with each object seemed to relate most strongly to the positioning of the objects within the interface, rather than with any characteristic of the objects themselves. Those objects that were located in the first set received the most tags, with the number of tags dwindling as the sets progressed. Here, sets were organized by institution and the Steve Project Team dismisses the phenomenon as the byproduct of taggers simply “trying out” the interface, something, which is likely true.

Set generation in OPAC 2.0 is based upon user input in the search form, and therefore the binding of resources into sets is based upon a characteristic in which the user already has some interest. Therefore it is difficult to assess the impact of sets on user tagging. What is recognized however from the Powerhouse Museum’s collected search narratives is that users, when searching across tags, tend to seek out specific objects from the resulting set, sometimes scrolling and clicking through several pages of results to do so. This potentially indicates that the notion of sets may have little relevance to tagging behavior.

6.5. RELEVANCE OF TERMS

A third concern that has been raised in regards to collaborative tagging in general, and more specifically regarding one element of system design is that of the relevance of the terms which are used as tags. One of the primary arguments against tagging is that, despite the understood flaws of the semantic gap, if taggers are left to tag at will with no restraints on terminology, the resulting metadata may be inappropriate, or misleading. In this scenario, collaborative tagging would actually impede accessibility, rather than boost it. The notion of having to extensively and systematically review terms poses an argument against the tagging system. Here however, both Steve and OPAC 2.0 can answer some questions.

The proof-of-concept test undertaken by Steve participants at the Metropolitan Museum of Art in the fall of 2005 posed two questions: First, were useful terms supplied by non-professionals? Second, was accessibility to tagged collections increased? The answer to both of these queries was resoundingly positive. In the test, collected tagging data was compared to terminology already existing in museum documentation, and specifically from the collections management system.

“Using a simple perl script (authored by Koven Smith) distinct tags per work were matched on a word-by-word basis (ignoring compound terms) to extracted words from the database records. (This simplest comparison would produce the largest number of matches). There were 3780 distinct terms in the 6679 terms collected. Of these distinct terms, 437 appeared in the Museum’s records, 3343 were new. The non-professionals made a contribution: 88% of the terms supplied by participants in the Proof of Concept Tests were not found in the basic museum descriptions.”²⁸

Furthermore, when these new terms were classified as “valid” or “invalid” according to a Subject Cataloging Committee comprised of professional catalogers, 77% were classified

²⁸ Trant, Proof of Concept, 2006

as “valid,” meaning they were considered appropriate to the work of art they were describing.

Though OPAC 2.0 lacks an analogous data set, it hints at a similar story. In the initial testing period of that project, 3,928 tags were submitted. In OPAC 2.0, the policing of tags is left to the users. Any user can delete any tag, at any time, removing the burden of review from the museum staff. Of the total number of tags, only 537 were deleted, or in other words, a validity rate of 86.4%. This number is only marginally higher than the percentage accepted by the (presumably more critical) Cataloging Committee. Incidences of commonality between these tags and existing museum documentation is unknown, however, as stated, OPAC 2.0’s design integrates tagging within a greater collections database environment. Therefore tags and museum documentation are understood to coexist side-by-side, and complement each other appropriately, even to the extent of repetition. The rapid increase in search traffic and the increased amount of time users spent on the website hints that the experiment is working as intended. Despite being supplied by non-professionals, despite coming from a nonstandard source, and despite being derived from personalized conceptions of the resource on the part of the user, it would seem that the user-supplied tags are useful, relevant, and effective.

6.6. MAKING THE USER-RESOURCE CONNECTION

What all of this seems to suggest more than anything is that the relationship between the user and the system is perhaps not as formative as one may initially suspect. Rather, it is the relationship between the user and the *resource* which is the strongest factor in tag formulation among museum patrons. But how then can museums understand, control, and foster this relationship? If the potential for collaborative tagging lay largely in the collections themselves, is there anything a museum can do to compel tagging, particularly since, as noted, the user has no part in resource creation? Whether conscious of this or not, both model tagging systems have developed methods, which are intended to exploit the innate user-resource relationship in a manner that benefits the museum.

Steve's approach to this dilemma is one borrowed from tagging systems outside of the museum domain. The key is the ability of users to register on the website, and maintain a personal log of works tagged. This enables users to essentially assemble their own personal digital collection of works in which they have a particular interest. This is essentially the nearest a museum can get to mirroring the type of connection Flickr users have to digital photographs they have taken, or of Last.Fm users to songs they have recently listened to. By participating in this function, the digital objects *belong* to the user. He or she is identified with them in some small way, and the possibility for forming user relationships or communities based on shared tastes is introduced. There is a practical advantage to integrating this type of experience in that many museums with digitized collections already offer similar functionality through their collections websites,

albeit without the tagging aspect.²⁹ If a collaborative tagging system were integrated with this type of program, which has already been shown to be popular, it might in the end facilitate a higher level of participation in the tagging system.

A problem with this approach however, to return to a nagging dilemma, is that one can gesticulate that the patrons who would be most likely to make use of such a feature would already be intimately familiar with the objects they are “collecting” in their personal galleries, and therefore would lose some of the spontaneity and insight that a relative novice might bring to the tagging table.

The Powerhouse Museum’s approach to the user-resource is somewhat different, and does much to circumvent the issue of users tagging only their favorite works. Rather than focusing upon the individual user in the user-resource relationship, OPAC 2.0’s developers have created a new feature more attuned to the unchanging nature of the resource. The experimental Bulk Tagger of the Powerhouse Museum, developed in the summer of 2007 and currently in its beta stage, presents the user with five randomly selected objects from the museum’s collections which have not yet been tagged at all.³⁰ Recognizing that only a fraction of the museum’s objects were being tagged, and perhaps also capitalizing on the notion that users might find themselves “on a roll” if presented with multiple sets of objects to tag at once, the Bulk Tagger represents a quick and easy way to give the user the most opportunities to view and tag objects in the shortest amount of time. It is currently unclear to what extent this sort of “object saturation” affects tag

²⁹ A popular and well-known version of this is available on the Metropolitan Museum of Art’s website, under the MyMetMuseum feature.

[<https://www.metmuseum.org/mymetmuseum/index.asp?HomePageLink=mymetmuseum>

³⁰ [<http://www.powerhousemuseum.com/dmsblog/experimental/bulktagger/index.php>]

production. It should be noted that this format is not entirely dissimilar from a game, which is one of the identified tagger motivations.

7. CONCLUSIONS

With a sound understanding that the collaborative tagging system is driven by a distinct set of relationships between the user, the resource, and the tagging system itself, one can begin to make assessments on the value of collaborative tagging relative to traditional cataloging. It has been set forth that collaborative tagging rectifies the issues that have been bequeathed to museums by years of imperfect and poorly adapted cataloging methods. It has also been suggested that collaborative tagging offers a means for museum patrons to wrest some level of control over collections from curators and museum staff, and to force museums to reevaluate their educational value. But how specifically does collaborative tagging accomplish these measures, and what are some other advantages, and even potential disadvantages?

It is clear that traditional cataloging of the type practiced consistently in museums for the last fifty years no longer meets the needs of museum patrons in the digital age. In order to be truly effective, museum metadata in the 21st century must speak in a voice which is universal and which cuts across lines of expertise, class, and age. Web 2.0 has given rise to a growing segment of the population that expects to have a functional role in the creation of content in all domains. Perhaps even more importantly, museum metadata must be assembled in a manner that is adaptive to any number of specific museum contexts, and it must be able to be created at a pace which keeps step with the forward progress of museum digitization projects. To presume that the same information that is suitable for an internal collections database will also work on a wall text, a website, or in a printed catalog is negligent on the part of the institution. While this level of description is certainly better than no description at all, the missions of museums and the

expectations of patrons demand something better. While some may suggest that simply increased attention on the part of the museum to cataloging is a solution, traditional cataloging by the methods already in place is burdened with flaws. These flaws are not specific to individual museums or catalogers, but rather are ingrained in the very processes of cataloging. The implementation of collaborative tagging of digital museum collections offers an alternative that promises to resolve these issues that plague cataloging and description, and which ultimately benefits the museum and the patron alike.

As defined earlier, one of the most persistent flaws of traditional cataloging in museums is the notion of standardization. It is the very uniqueness of museum objects that make them valuable, and therefore attempting to identify and isolate threads of relevant information across entire museum collections is difficult. It is increasingly clear that the inherent value of museum objects cannot be captured within a standardized descriptive rubric, however, it is just such a rubric upon which the entire premise of traditional cataloging is based. Collaborative tagging circumvents this issue by allowing users to formulate descriptive tags based on the aspects of a museum collection object that they personally find to be important, thought provoking, or interesting. A patron participating in a tagging system need not necessarily have knowledge of an artist's name or origin to notice something about a work that might be useful to other like-minded individuals. The standard descriptive fields of museum metadata such as title, nation of origin, and date overlook what draws many individuals to museum objects in the first place, their aesthetic, ethereal, and intangible qualities. Though these qualities frequently

manifest themselves in a context that is entirely personal to the user, collaborative tagging recognizes this and offers an outlet.

Even if one were to attempt to classify large groups of museum objects by artist, origin, media, or other trait, the lack of a well-developed standardized vocabulary poses a second obstacle. Without such an all-encompassing tool, the work of individual catalogers working in museums would be devalued by slight terminological differences across institutions. The notion of sharing information is impeded by differing personal, local, and regional opinions on such blurry fields as genre, or even date. Projects such as the Getty Vocabularies Project show some potential for resolving these issues, especially as they pertain to proper names. But again, given the uniqueness of museum collections such a vocabulary would be difficult to implement, and even more difficult to maintain, and would only be relevant to a specific subset of institutions.

Even if a far-reaching standardized vocabulary project could be undertaken with some degree of success, there is a deeper sociological objection to the controlled vocabulary, that being the underlying issue of the “semantic gap.” The terminology of the museum professional and the art historian is (somewhat understandably) highly specialized. If the ultimate goal of describing collections through cataloging is to increase accessibility to museum users, employing a niche vocabulary comprised of terms that may be relatively unknown or unclear to the layperson is counterproductive.

That collaborative tagging allows users to draw their terminology from any uncontrolled source is a tacit acknowledgment that while standardized vocabularies may be useful to a degree, they enable the perpetuation of the “semantic gap.” There is a significant byproduct of this ontological freedom that can be seen in tagging projects of

all domains, which is the formulation of a new folksonomy. The folksonomy is essentially a locally developed set of terms, formulated organically through consensus and with a greater level of participation from the communities who make use of it. This is in direct opposition to a controlled vocabulary, in which terminology is chosen by a small subset of individuals and is sometimes developed in a manner that seems contrary to common usage or understanding. Though it behaves in much the same manner as a controlled vocabulary, the folksonomy greatly increases accessibility without succumbing to the perils of the “semantic gap.” Its focus on usability rather than standardization enables greater freedom for adjustment and adaptation to respond to shifts in vernacular across communities.

Collaborative tagging, with its tied contributions to folksonomy development, need not ring the death knell of controlled vocabularies altogether. Rather, it should be viewed as a tool that can be used for their improvement. As data from the Steve project shows, it is not uncommon for terms from standardized vocabularies to appear in user-supplied tags.³¹ Is it not possible then that collaborative tagging could be used as a comparative measure for gauging the ultimate effectiveness of standardized vocabularies, and provide a guide for their ongoing improvement? Without question, one of the underlying themes coupled with debate about collaborative tagging as a form of cataloging is that of tradition versus innovation. The common implication is that new developments such as collaborative tagging exist in isolation, and will replace older methods entirely. The ultimate goal, however, of any museum’s approach to technology

³¹ Trant, Proof of Concept, 2006

should be fulfilling its mission more effectively. If this entails a mixture of the old and new, so be it.

Another issue related to the cataloging of collections that the “in-hand” method fails to address is the question of determining what is “important” about museum objects. The complexities of museum objects, particularly artworks, demand some form of recognition that their innate significance is not uniform across museum patrons or even across institutions. Each museum object is situated within a very specific context, which, depending on the artwork or artifact, can be very different. For any single object the point of significance may be the creator, the subject, the media, the date it was produced, or any combination of these. As shown, collaborative tagging places no restrictions on the nature of the terms used as tags. Individual taggers are expected to focus upon the aspects of an object that they are familiar with, and that therefore are important to them. With many taggers acting in concert, description will ultimately be broader, without necessarily sacrificing specificity or accuracy.

Collaborative tagging also offers relief from some of the practical issues which plague cataloging as related to the use of staff time and expense. Cataloging is a time-consuming endeavor that requires trained staff. These staff must be knowledgeable in cataloging methods as well as in the subject areas that pertain to collections, a combination that in many cases is difficult to find. Cataloging is then often undertaken by other staff members, for whom it is something of an afterthought among the day-to-day operations of the museum. It is for this reason that many museums sit atop backlogs of accessions that lack the necessary descriptive information to make them accessible to a desirous public. As a form of cataloging which that is based entirely upon the willing

participation of a large patron pool, the speed of tagging is hindered only by the ability of the museum to involve users.

Finally, collaborative tagging has a broader and somewhat more abstract value, and that is the way in which it challenges traditional concepts of who is the authority of knowledge and learning in museums. The notion that the educational value of collections is a one-way street in which curators and museum professionals hold the sole authority over learning is preposterous. Part of the reluctance that many potential patrons hold towards museums is based on a shared (and not always misplaced) conception of them as indomitable fortresses of learning. Thus it is that patrons travel through exhibitions as sponges, listlessly soaking in information until they are ready to burst. Collaborative tagging creates a personal relationship between the museum, the patron, and the collections, one that transcends such staid traditional learning paradigms as teacher-student.

Though it has been shown that collaborative tagging addresses the salient issues of traditional cataloging in museums and offers an alternative that is overwhelmingly cost-effective, it would be remiss not to mention some of the potential objections to the practice itself, many of which have a solid grounding. First and foremost among these is the uncertainty of collaborative tagging. Though it has precedents in fields outside of the museum domain, the entire concept remains infantile in terms of its development and understanding. Any predictions of its long-term effects and benefits are, at best, conjectural. Even the data collected on a short-term basis by projects such as Steve and OPAC 2.0 is generally oriented around ascertaining the optimal technical setup for such a project, not for determining its affect on the use of digital museum collections and its

impact on visitors. Put simply, while one can say with some certainty that collaborative tagging works as a tool for describing collections, it is yet to be seen how it fits into the greater mission of increasing accessibility to collections, dispersing information, and improving the relationship between the museum and current patrons, while at the same time attracting new groups.

A second concern commonly voiced regarding collaborative tagging pertains to the ultimate accuracy of the resulting metadata. With little restriction on contributors, is it not possible for the efficacy of a collaborative tagging system to be marred by improper and inaccurate tagging, even taken to the extreme of sabotage and abuse? As many point out, museum professionals are trained and equipped to approach metadata formulation with a certain level of seriousness and the knowledge that description must be formulated with longevity in mind. As a response to these fears, most tagging systems maintain a certain level of supervision in which supplied tags are somehow checked before being fully implemented into the system, and those which contain misspellings, or which are deemed otherwise inappropriate are removed. Though this question of accuracy is, above all, a matter of opinion and preference in which a strong case can be made for either side, the data of the Steve project suggests that such fears may be unfounded. Of the tags supplied during the beta tests, only a small fraction of these were deemed unsuitable, most due to misspellings and other typos.³²

It can be argued that these and most other objections are essentially the byproduct of resistance on the part of museum professionals to change and innovation, a common malady, which is not entirely surprising given the dedication of the profession to concepts

³² Trant, Proof of Concept, 2006

such as conservation, preservation, longevity, and order. Nonetheless museums in the 21st century are at a crossroads where they are forced to balance their commitment to their collections with their service to their patrons. Collaborative tagging, when placed side by side with the needs and missions of museums today, is hardly contrary to the ideals upon which institutions are built. To shed the image of dark storehouses of art and artifacts is to acknowledge and embrace the concept of collections as the building blocks of education, understanding, and personal growth. By placing these collections, or digital representations thereof, in the hands of as many people as possible is to act on this idea in a manner that is entirely responsible, and ultimately rewarding for all involved.

Works Consulted

Bearman, David and Jennifer Trant. "Social Terminology Enhancement through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections." *D-Lib Magazine*. Volume 11, Number 5 (September 2005).

Chan, S., Tagging and Searching – Serendipity and museum collection databases. In J. Trant and D. Bearman (eds). *Museums and the Web 2007: Proceedings*. Toronto.

Elings, Mary W. and Gunter Waibel. "Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives, and Museums." *First Monday*. Volume 12, Number 3 (March 2007).

Harpring, Patricia. "How Forcible are Right Words! *: Overview of Applications and Interfaces Incorporating the Getty Vocabularies." in *Proceedings of Museums and the Web Conference: 1999*.

Institute of Museum and Library Services. "Status of Technology and Digitization in the Nation's Museums and Libraries 2002 Report." 2002.

Lin, Xia, Joan E. Beaudoin, Yen Bui, and Kaushal Desai. "Exploring Characteristics of Social Classification." in *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop: November 2006*.

Marlow, C., Namaan, M., Boyd, D., and Davis, M., "HT06, Tagging Paper, Taxonomy, Academic Article, ToRead", *Proceedings of the 17th Conference on Hypertext and Hypermedia: 2006*.

Mesnagi, Cedric, and Jazzari, Mehdi. "Specifying the Collaborative Tagging System." *Proceedings of the Semantic Authoring and Annotation Workshop: 2006*.

Tonkin, Emma. "Searching the Long Tail: Hidden Structure in Social Tagging." *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop: November 2006*.

Trant, Jennifer, et al., "The eye of the beholder: steve.museum and social tagging of museum collections," in *International Cultural Heritage Informatics Meeting (ICHIM07): Proceedings*, J. Trant and D. Bearman (eds). Toronto: Archives & Museum Informatics. 2007.

_____. "Understanding Searches of an On-Line Contemporary Art Museum Catalog." *Archives and Museum Informatics*. Fall 2006.

_____. "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept." Draft for *New Review of Hypermedia and Multimedia*: 2006 May 31.

van Hooland, Seth. "Spectator Becomes Annotator: Possibilities Offered by User-Generated Metadata for Image Databases." in *Proceedings of the CILIP Cataloguing & Indexing Group Annual Conference*, Norwich (United Kingdom) 2006.

Weinberger, David. "Tagging and Why it Matters." 2005.

Wyman, B., et al., Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums, in J. Trant and D. Bearman (eds.). *Museums and the Web 2006: Proceedings*, Toronto.