

Social Metadata for Libraries, Archives and Museums

Part 2: Survey Analysis

Karen Smith-Yoshimura
OCLC Research

Carol Jean Godby
OCLC Research

Helice Koffler
University of Washington

Ken Varnum
University of Michigan

Elizabeth Yakel
University of Michigan



A publication of OCLC Research

Social Metadata for Libraries, Archives and Museums. Part 2: Survey Analysis
Karen Smith-Yoshimura, Carol Jean Godby, Helice Koffler, Ken Varnum and
Elizabeth Yakel, for OCLC Research

© 2011 OCLC Online Computer Library Center, Inc.

Reuse of this document is permitted as long as it is consistent with the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 (USA) license (CC-BY-NC-SA): <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

December 2011

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 1-55653-425-6 (978-1-55653-425-6)

OCLC (WorldCat): 712604667

Please direct correspondence to:

Karen Smith-Yoshimura

Program Officer

smithyok@oclc.org

Suggested citation:

Smith-Yoshimura, Karen, Carol Jean Godby, Helice Koffler, Ken Varnum and Elizabeth Yakel.
2011. *Social Metadata for Libraries, Archives and Museums. Part 2: Survey Analysis*. Dublin,
Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-03.pdf>.

Table of Contents

Acknowledgements	7
Overview	8
Trends and Themes	9
Methodology	10
Reference	11
Background of Social Metadata Survey Respondents	12
Objectives	14
Content	14
Audiences	16
Social Media and User Contribution Features and Use by Ken Varnum	18
Controlled Vocabularies	20
Tagging	22
Uses of Social Metadata and User-generated Content	23
Tools and Functionality by Karen Smith-Yoshimura	27
Open-Source Software	28
Functionality	29
Reference	30
Moderation by Carol Jean Godby	31
Staffing and Site Management by Helice Koffler and Karen Smith-Yoshimura	34
Site Management	37
Policies	40
Examples of Different Policies	42
Policies that Extend Existing Institutional Policies	43

Examples of New Policies	44
Policies that Combine Extensions to Existing Policy and New Policy	45
References.....	46
Measuring Success by Elizabeth Yakel	49
Appendix A: Sites that Responded to Social Metadata Survey at a Glance	54
Appendix B: Social Media and User Contributions Offered at a Glance	57
Appendix C: Social Metadata Survey Questions	60

Tables

Table 1: Definitions of site types	13
Table 2: Key objectives for offering social media	14
Table 3: Types of content on site	15
Table 4: Communities respondents' sites are engaging with	16
Table 5: Evaluations respondents do with site participants.....	17
Table 6: Uses of tagging	23
Table 7: Uses of social metadata	24
Table 8: Uses of user-generated content	24
Table 9: Types of tools used	27
Table 10: How search results are ranked	30
Table 11: Moderation	32
Table 12: Tracking users	32
Table 13: Spam	33
Table 13.5: User abuse	33
Table 14: Staff who play a role in site management.....	35
Table 15: Percentage of time spent on different site maintenance activities	36
Table 16: Site management.....	40
Table 17: Types of policies or guidelines implemented	41
Table 18: Types of activities measured	51

At a Glance: Sites that Responded to Social Metadata Survey
spreadsheet: <http://www.oclc.org/research/publications/library/2011/2011-03a.xlsx>

Figures

Figure 1: Countries represented in sites that responded to Social Metadata Survey	12
Figure 2: How long social media features have been offered by respondents	13
Figure 3: Sources of content on respondents' sites	15
Figure 4: Social media and user contribution features offered by respondents.....	19
Figure 5: Controlled vocabularies used in respondents' sites to describe content	21
Figure 6: How respondents use social networking content for desc. of assets or collections..	26
Figure 7: Roles staff serve on site	34
Figure 8: Staff time committed to creation and maintenance of the site	36
Figure 9: Frequency with which new content is added to site	38
Figure 10: Success criteria	50
Figure 11: Data capture methods	51
Figure 12: Frequency of unique visitors per month	52
Figure 13: Visitors contributing content per month	53

Acknowledgements

This report was created in collaboration with the following members of the RLG Partners Social Metadata Working Group (for short biographies, see “Who Are We?” at www.oclc.org/research/activities/aggregating/group.htm).

Drew Bourn, Stanford University
Douglas Campbell, National Library of New Zealand
Kevin Clair, Pennsylvania State University
Christopher Cronin, University of Chicago
Christine DeZelar-Tiedman, University of Minnesota
Mary Elings, University of California, Berkeley
Steven Galbraith, Folger Shakespeare Library
Carol Jean Godby, OCLC
Cheryl Gowing, University of Miami
Rose Holley, National Library of Australia
Rebekah Irwin, Yale University
Lesley Kadish, Minnesota Historical Society
Helice Koffler, University of Washington
Daniel Lovins, Yale University
John Lowery, British Library
John MacColl, OCLC
Marja Musson, International Institute of Social History (Amsterdam)
Henry Raine, New-York Historical Society
Cyndi Shein, Getty Research Institute
Ken Varnum, University of Michigan
Melanie Wacker, Columbia University
Kayla Willey, Brigham Young University
Elizabeth Yakel, University of Michigan, School of Information

Overview

The cultural heritage organizations in the RLG Partnership were eager to take advantage of user contributions to enrich the descriptive metadata created by libraries, archives and museums (LAMs)¹ and expand their reach into user communities. Enriching LAM metadata improves the quality and relevancy of users' search results and helps people to understand and to evaluate the content better. User contributions can also augment and provide additional context to LAM resources. In 2009–2010, a 21-member RLG Partner Social Metadata Working Group from five countries investigated how to take full advantage of the array of potential user contributions that would improve and deepen their sites' user experiences. The working group considered issues related to assessment, content, policies, technology, and vocabularies.

In our first report, *Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews* (Smith-Yoshimura and Shein 2011), we reviewed 76 sites of most relevance to libraries, archives and museums that supported such social media features as tagging, comments, reviews, images, videos, ratings, recommendations, lists, links to related articles, etc. We refer to information created by these features as social metadata. In this, our second report, we analyzed the results from a survey of site managers conducted in October–November 2009. We received 42 responses to our survey; 40% of the respondents were from outside the U.S., including 10% from the United Kingdom and 24% from Australia and New Zealand. The sites that responded originate from academic libraries and archives, national libraries or archives, non-profit organizations not affiliated with any institution, museums, historical societies, consortia, other cultural institutions, public libraries, plus one botanical garden and one special library. The survey focused on the motivations for creating a site, moderation policies, staffing and site management, technologies used, and criteria for assessing success.

The third report provides recommendations on social metadata features most relevant to libraries, archives and museums and factors contributing to success. We looked at barriers that need to be overcome, and the lessons we learned from the site reviews, survey responses, interviews, and the literature. This report contains a reading list of resources referenced

1. Some countries use the term "GLAM" for galleries, libraries, archives, and museums.

during our research. We hope that our recommendations will enable cultural heritage institutions to leverage users' enthusiasm while enhancing their own resource descriptions and extending their reach to new communities.

Trends and Themes

- Most sites have been offering social media features for a short time—more than 70% had been offering social media features for two years or less. The respondents represent active and current sites; 83% of respondents add new content at least monthly.
- Building user communities and increasing traffic to expose the site's content are key objectives.
- Most respondents manage their own sites rather than use hosted services, perhaps reflecting that more respondents come from larger organizations than smaller ones that would more likely use hosted services.
- Sites are increasingly multi-media; although still images and text predominate among the responding sites, more than a third also offer moving images and audio. Archives are a predominant source of content.
- The general public is the target audience for almost all responding sites. Academics are a key audience, especially for library and archive sites.
- Usability testing tends to be done later in a site's life cycle rather than as part of the development stage.
- Comments, tagging, and RSS are the most common social media features offered. Only half of the sites using reviews also used ratings.
- More than half of the survey respondents use a controlled vocabulary on their sites.
- Only half of respondents indicated that they show users tags already in the system. A third combine user-contributed tags with their own controlled terms.
- A minority of survey respondents are concerned about the way the site's content is used or repurposed outside the site.
- Most respondents index user-supplied metadata; most user-supplied content is searchable. More than half correct existing metadata as the result of user

contributions. However, a minority incorporates metadata into their own description workflows and incorporates user-contributed content into their own sites.

- More than half of the sites use a combination of open-source software and software developed internally.
- A majority of sites moderate user contributions, and half edit user contributions before they are posted. Spam and abusive user behavior are sporadic and easily managed.
- The majority of staff responsible for site management seem to be drawn from the information technology departments and as a part-time responsibility of professional staff (archivist, curator, or librarian.) Mature sites spend more time on adding new content and moderation than newer sites.
- A number of respondents are integrating their sites into institution's production services rather than being dependent on external or temporary funding sources.
- The majority of sites have policies concerned with appropriate behavior, rights to edit or remove content and safeguarding privacy. Policies vary greatly in both depth and scope, but reflect the shared concerns of LAMS that are opening their content to social interaction. LAMs are making efforts to maintain a safe environment for users, with particular attention to under-aged users, and upholding professional ethics and laws to provide equal access and protect intellectual property rights.
- The vast majority of respondents consider their sites to be successful, regardless of the type of institution (library, archive, museum), whether the site is managed locally or uses a hosted service, or the amount of interaction on the site.
- Engaging new or existing audiences is used as success criteria more frequently than adding new content or gathering metadata about existing content.
- The survey results indicate that engagement is best measured by quality, not quantity.

Methodology

The working group divided itself into subgroups focusing on assessment, content, policies, technology, and vocabularies. These divisions arose from the site reviews we conducted for our first report and the literature we consulted. Each group developed questions to include in the survey, which were then put together by a survey design subgroup and formatted in SurveyMonkey. The survey included logic questions, where some questions were prompted

only if a person had responded in a specific way to a previous question. Volunteers from some of the sites that we had reviewed pilot-tested the survey before it was distributed. A link to the survey was also posted on several e-mail discussion lists. The primary population for the survey was the 76 sites we had identified for our first report. Invitational e-mails were sent to a contact from each of these sites. Of the 42 responses, 26 came via the e-mail invitation and 16 from the web link. All but six of the sites that responded to the survey are represented in our first report, making the response rate for that sample 47%.

Survey responses were downloaded into Excel. Duplicate responses were removed and responses that had similar meanings were normalized. Working group members analyzed the responses in conference calls and used Basecamp, a web-based project management and collaboration tool, to share their analysis and documents; volunteers drafted sections of this report. The list of the responding sites with some of their characteristics is in the “Sites that responded to the social metadata survey at a glance” (see appendices A and B).

The survey questions are included as appendix C to this report.

Reference

Smith-Yoshimura, Karen and Cyndi Shein. 2011. *Social Metadata for Libraries, Archives and Museums. Part 1: Site Reviews*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-02.pdf>.

Background of Social Metadata Survey Respondents

The site managers who responded to the survey come from seven countries. Responses from U.S. site managers represent the majority (60%). Eight responses came from Australia, four from the United Kingdom, two from New Zealand and one each from Germany, the Netherlands and Spain (asterisked as "Other" in the figure below).

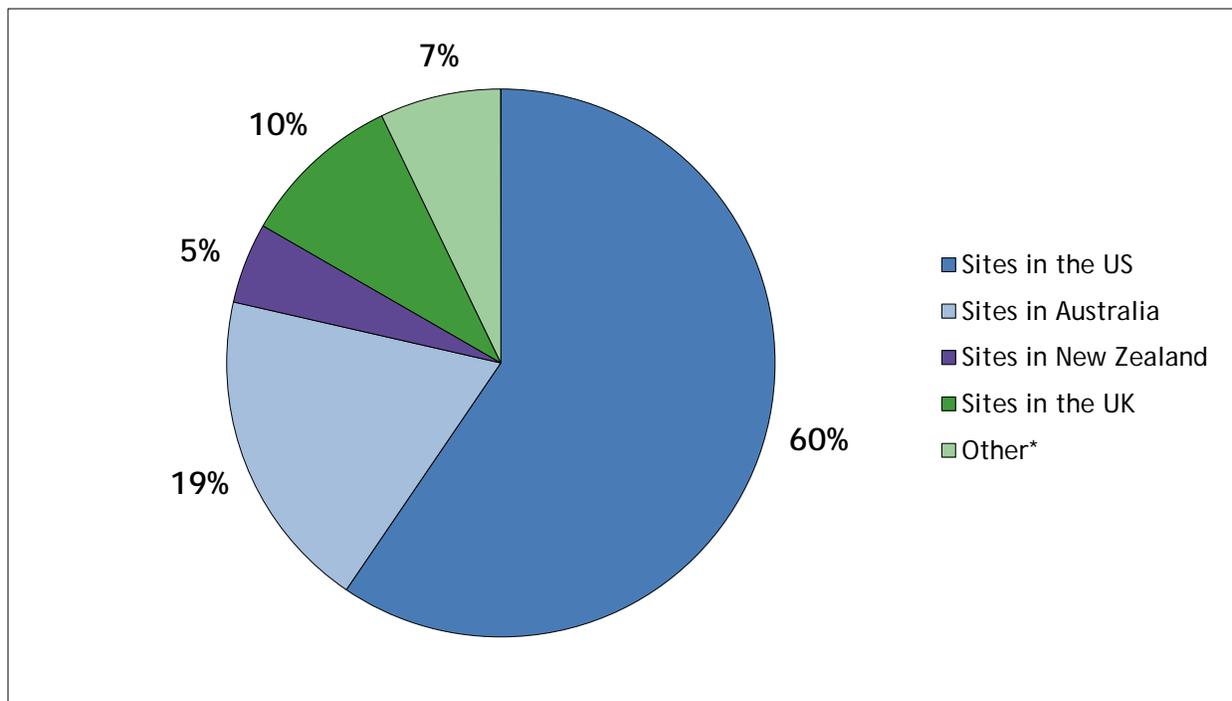


Figure 1: Countries represented in sites that responded to Social Metadata Survey (n=42)

We also characterized the type of site that responded using the same definitions as in our site reviews. Some sites had multiple types assigned.

Table 1: Definitions of site types

Type	Definition	Number of Respondents
Archive	Site is hosted by an archive and/or provides access to primary source materials.	19
Community	Content provided primarily by individual volunteers. Some organizational sites may also be tagged "Community" if the content provided by individuals is significant (e.g. editing text, uploading images, etc.) Simply the availability of social media features, such as tagging of images, does not qualify a site for inclusion in this category.	10
Discipline	The focus is entirely discipline-based. The site was not created to show-case the collection of specific institution(s), but to facilitate communication and information sharing in a specific discipline. May be hosted by any type of organization.	5
Library	Site is hosted by a library and/or provides access to library materials.	24
Museum	Site is hosted by a museum and/or provides access to museum resources	7

A large number of organizations have been offering social media features for only a short time—more than 70% have been offering social media features for two years or less. Four sites were not even public yet at the time of the survey. On the other hand, eight sites (19%) have been offering social media features for four years or more.

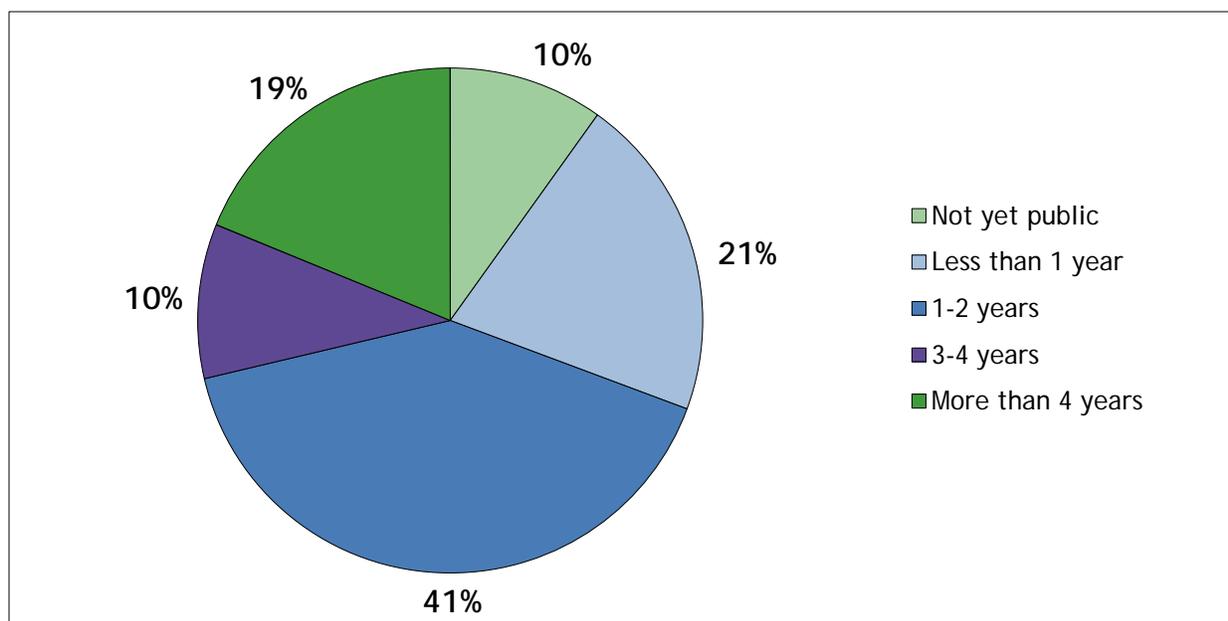


Figure 2: How long social media features have been offered by respondents (n=42)

Most of the survey respondents managed their own sites—79%. The working group was surprised by this, given the draw on staff resources. Ten sites (24%) use a hosted service; one site both manages its own site and uses a hosted service. Larger organizations tend to manage their own sites as well as expose their collections in hosted sites like Flickr. Small organizations were not as well represented in the survey population; we would expect them to use hosted sites rather than manage their own.

Objectives

Site managers had multiple objectives for offering social media on their sites or using the social media features on hosted sites.

Table 2: Key objectives for offering social media (multiple responses allowed) (n=42)

Objective	Number of Responses	Percentage
Build user community	34	81%
Increase traffic and access to our content	33	79%
Enhance description	25	60%
Build collection	11	26%
Other (please specify)	4	10%

We correlated the responses by type of organization. Building community is a key interest across all types. Academic libraries and archives tend to be more interested in increasing traffic to their sites, providing better access to their content, and enhancing description. They are less interested in acquiring additional content from other sources. National- and state-level institutions are more likely to seek additions to their collections. The other responses came from museums interested in inspiring visitors and getting them more involved with exhibits and museum activities.

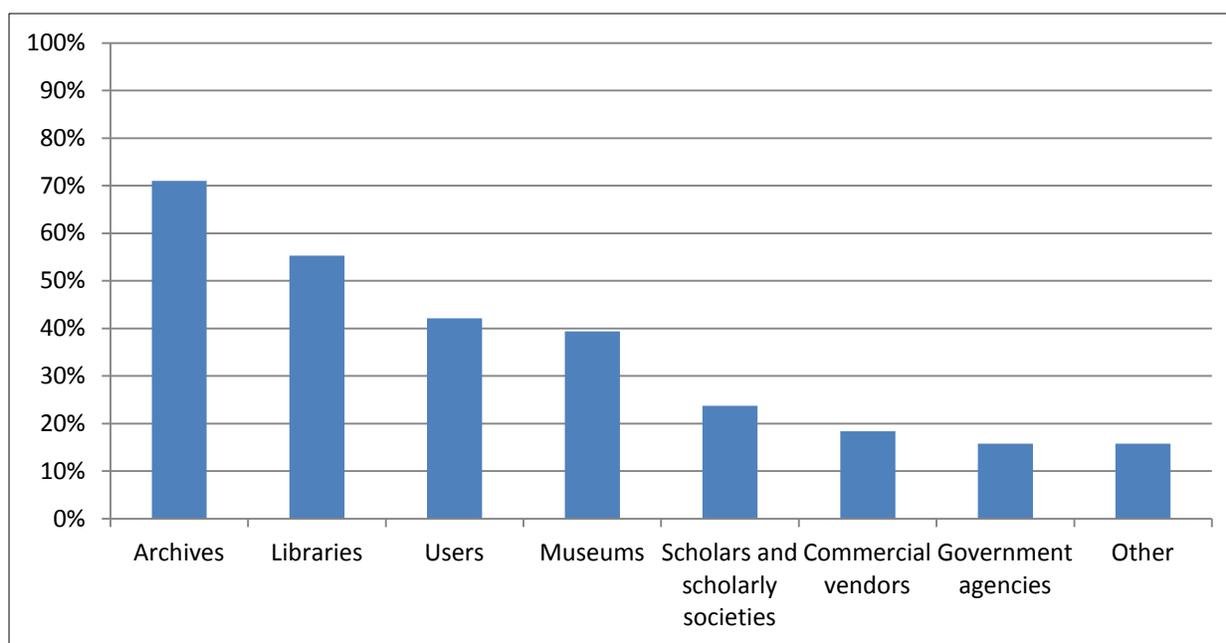
Content

Almost all sites offered still images (91%) followed closely by text (86%). Many of the respondents offered moving images (43%) and audio (38%). Among the other types of content offered: links, bibliographic references, interactive maps, calendars, live content via Twitter, tabular data, flash games and other interactive applications and virtual page-turning books.

Table 3: Types of content on site (multiple responses allowed) (n=42)

Type of Content	Number of Responses	Percentage
Still images	38	91%
Text	36	86%
Moving images	18	43%
Audio	16	38%
Other (please specify)	7	17%

We were interested in learning about the source of materials libraries, archives and museums were making available for user interaction. Sources of the content reflect the types of organizations that responded to the survey. The four main sources were archives (71%), libraries (55%), users (42%), and museums (39%). Most who responded “archives” also responded “libraries.” At the low end of the scale were commercial vendors (18%) and government agencies (16%). Commercial vendors’ content is likely to be more difficult to incorporate into social media sites because of licensing or other contractual restrictions. We surmise that government agencies more likely place restrictions on user interaction. The “other” category included galleries, RSS feeds and tweets, social networking sites such as YouTube, Vimeo, Flickr, newspapers, public broadcasters and Creative Commons licensed journals.

**Figure 3: Sources of content on respondents' sites (n=38)**

Audiences

The general public is the most common audience targeted by the site respondents with 92%. Academics are target audiences: college and university students (73%), academic faculty, teachers, and independent scholars (62%-68%). The academic libraries and archives among the survey respondents particularly targeted these audiences. Professionals were cited by 51% of the respondents. Hobbyists did not rank very high (43% of the respondents); they were less likely the target of the libraries and archives sites. All collections that use Flickr target hobbyists.

Table 4: Communities respondents' sites are engaging with (multiple responses allowed) n =37

Community	Number of Responses	Percentage
General public	34	92%
Students: College and university	27	73%
Academic faculty	25	68%
Independent scholars	24	65%
Teachers	23	62%
Professionals	19	51%
Hobbyists	16	43%
Students: between ages 14 and 18	14	38%
Students: between ages 6 and 13	9	24%

70% of the respondents feel that they have encouraged user communities by allowing user-contributed content. Fifteen of these 25 respondents also selected "build user community" as one of their key objectives and indicated that "engaging new audiences" was one of the criteria for measuring success for their sites.

Site managers tend to be more reactive than proactive in conducting evaluations with site participants. The most popular method is relying on a feedback/comment/suggestion box (67%). The other options provided, all proactive, received fewer responses, with usability testing ranking fairly high (36%). Two respondents said that usability testing was in the planning stage, which would then surpass those who do surveys (39%). Ten sites (28%) did no evaluations; one noted that their site was driven entirely by volunteers participating and the site wouldn't exist without them. Another reported that the feedback features were not active yet. The other evaluation reported was analysis of content categories of tags, comments, notes and tallying statistics for views, favorites, tags, comments, and notes.

**Table 5: Evaluations respondents do with site participants
(multiple responses allowed) (n=36)**

Type of Evaluation	Number of Responses	Percentage
Feedback/comment/suggestion box	24	67%
Surveys	14	39%
Usability testing	13	36%
None	10	28%
Interviews	8	22%
Focus groups	6	17%
Other	1	3%

We correlated the usability testing done by sites with how long the site had been offering social media features. Older sites had done the most usability testing; sites that had been around for 3-4 years had the most extensive testing. Usability testing had been done by only one site that was not yet public. Usability testing was done rather late in the site's life cycle rather than in the development stage. Surveys are also generally done later in the sites' life cycles. We speculated that these arise when something is not working or when priorities need to be set.

Social Media and User Contribution Features and Use

Ken Varnum

We presented a list of nineteen social media and user contribution features and asked respondents to select the ones they offered their patrons, with an option to describe a feature not listed. The top three features used by the 39 site managers who responded were comments (82%), tagging (67%), and RSS feeds (54%). RSS feeds may have come up so frequently because they are often an embedded feature of many open source and off-the-shelf software packages. Annotations (37%), upload materials (31%), user profiles (28%), user-contributed images (26%), bookmarks (21%), reviews (21%), and ratings (21%) made up the midrange. Four features were each supported by five sites (13%): edit text, user awareness (who else is logged on), form sub-groups, and user recommendations.

At the bottom of the scale are collaborative filtering and synchronous chat, with one response each (3%). We theorize that these are technologically advanced functions that may not be easily implemented. We also believe that synchronous chat will increase in the near future as social media tools such as Twitter and Facebook drive this form of interaction deeper into the public's toolkit. The "other features" offered were tweets, forums or forum posts, create blog entries and suggesting additions to the metadata fields used in the search index.

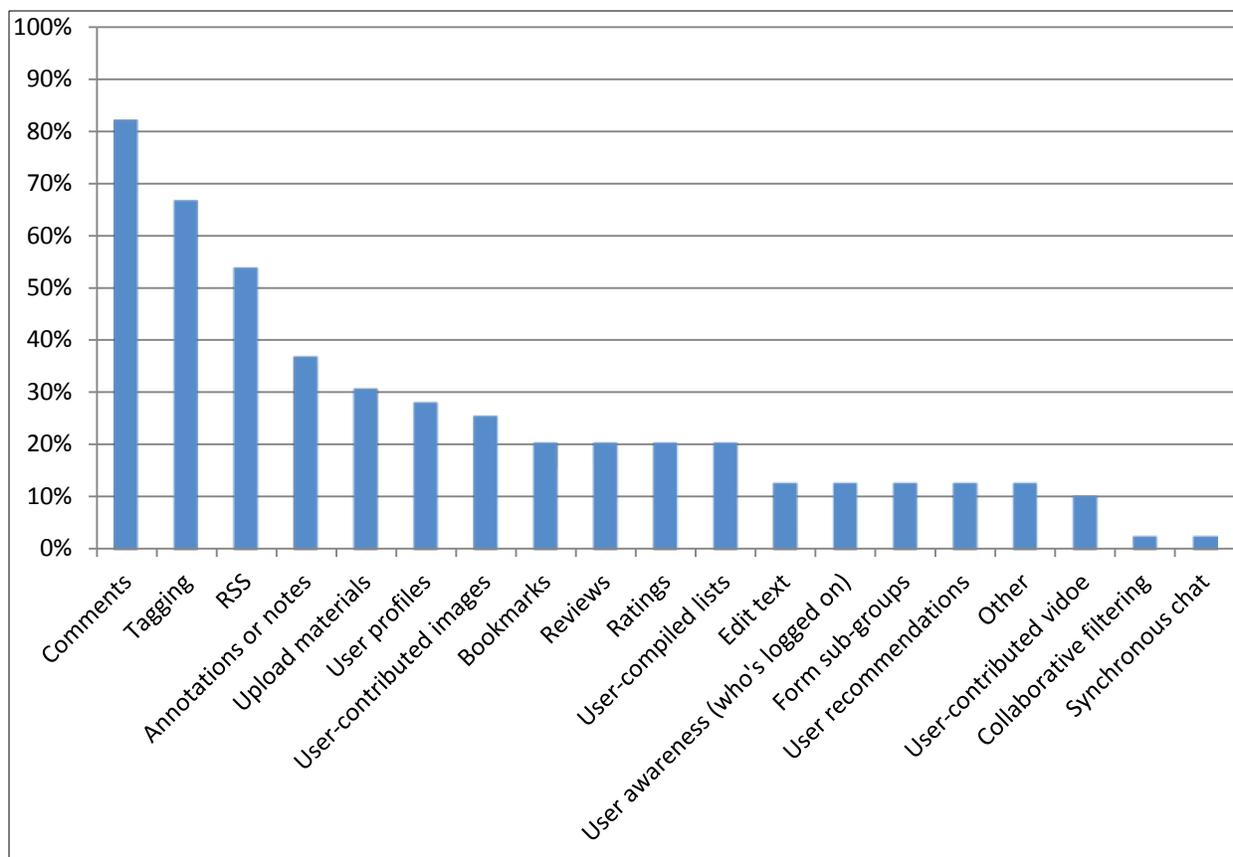


Figure 4: Social media and user contribution features offered by respondents (n=39)

We explored the correlation of features offered by sites using filters on the [Sites that Responded to the Social Metadata Survey at a Glance](#) spreadsheet. The 32 sites using “Comments” accounted for:

- 88% (7 of 8 responses) of the sites using bookmarks
- 83% (10/12) of sites uploading materials
- 81% (17/21) of sites using RSS
- 80% (8/10) of sites using user-contributed images
- 79% (11/14) of sites using annotations
- 77% (20/26) of sites using tagging.
- Of the 20 sites that use both tagging and comments, 65% (13/20) also use RSS.
- Of the 14 sites using annotations, 86% (12/14) also use tagging.

- Although eight sites reported using reviews or ratings, only half (4/8) of the sites using reviews also used ratings.

We reviewed the reasons respondents gave for choosing the social media features offered by their sites. The answers reflected the varied nature of the sites themselves. The sites using hosted services such as Flickr automatically received the other features Flickr supports: comments, tagging, and RSS feeds. According to one respondent, "Flickr's pre-existing, feature-rich set of social interactions and the fact that a highly involved community of users had begun to coalesce around Commons' images were certainly motivating factors." In a similar way, institutions that picked a particular software package to meet their primary need often ended up with additional social media features simply because those features were also provided, not because they met a pre-defined need.

Finally, some respondents reported not being able to use a desired feature because of policy or legal restrictions within the institution. One respondent noted, "As a government wiki we have to restrict the freedoms offered in non-government wikis; it would have been more interesting to offer further features had we been allowed."

Many respondents reported wanting to tailor the features they offered as much as they could to the specific audiences being served. One respondent wrote, "Each UGC [user-generated content] type was chosen with the intent to allow the user to interact with bibliographic records in ways that would help them better manage their material use and ways that would help them share information with others. We followed the example of public websites and considered the UGC types that typically had the most and quickest up take and provided the largest opportunity for re-use."

We were uncertain as to the impact the rapidly changing social media landscape might have in the near future. RSS has become more like plumbing than a true social media tool—it's a way to get data from place to place, but offers little interaction. For some audiences, Twitter is taking the place of RSS.

Controlled Vocabularies

We asked what controlled vocabularies, if any, people used in their sites. More people skipped this question than other survey questions, with 22 respondents (52%) saying they used a controlled vocabulary, notable given the intellectual overhead required to assign controlled vocabularies. Multiple responses were allowed.

Of those who report using controlled vocabularies, more than half use either the Library of Congress Subject Headings (LCSH) (64%) or a locally-developed thesaurus (64%). Of those that

use LCSH, only two responded that they use LCSH exclusively. LCSH is often used in combination with the Art and Architecture Thesaurus (AAT). Many of the sites that shared images used AAT (41%). This thesaurus is also used in archives for document types. LCSH is not used as much outside the Anglo-speaking world.

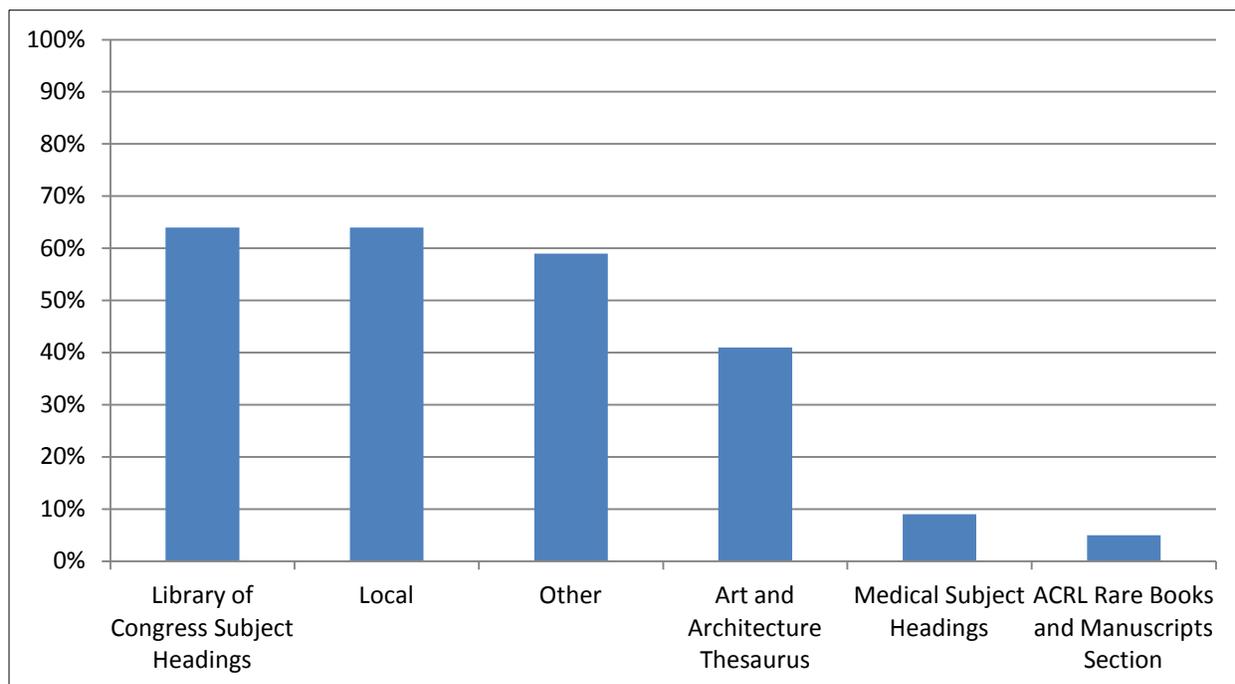


Figure 5: Controlled vocabularies used in respondents' sites to describe content (n=22)

The thirteen "other" responses:

- Australian Pictorial Thesaurus
- Audio Visual Materials Thesaurus for Tobacco Control Archives
- We use strictly controlled headings and extensive cross references for subjects, corporate bodies, places and personal names.
- Multilingual authorities are being built for what we use LCSH (English), the authority file of the National Library of Spain, authorities locally generated, and Virtual International Authority File (VIAF).
- GTAA (thesaurus of the Netherlands Institute for Sound and Vision) and Cornetto (database which contains the bulk of all official Dutch words)

- Library of Congress Thesaurus for Graphic Materials
- Custom thesaurus for multi-parent hierarchy of instrumental classifications
- We use multiple thesauri for Digital Gallery subject headings, which are parsed during uploading via a simple algorithm into more tag-like terms for Flickr
- Māori Subject Headings
- Multiple vocabularies for photographs across the institution
- U.S. Geological Survey

Tagging

Tagging was the second most common feature offered by sites. Of the 26 who indicated that they supported tagging, 24 completed a follow-up question on how they used tagging. Only half of respondents indicated that they show users tags already in the system. This seems counterintuitive, as we thought one of the main roles of tagging would be for users to see their tags, and those of other users, associated with the item. In our first report, we reviewed *Waisda?*, an example of an intuitive use of tagging. The site makes a game out of tagging digital objects, pitting two anonymous players against each other, tagging the same item simultaneously until they agree on a common tag.

Conversely, one third combine user-contributed tags with their own controlled terms. One quarter provide a list of suggested terms when users tag a resource. Respondents from two sites indicated that they offer users the opportunity to select from lists of controlled vocabulary to apply to images; this could reduce new terms suggested by users, tagging's primary purpose.

Table 6: Uses of tagging (multiple responses allowed) (n=24)

Use	Number of Responses	Percentage
Show users existing tags in the system	12	50%
Combine user-contributed tags with our own controlled terms	8	33%
Other	7	29%
Provide a list of suggested terms when users tag a resource	6	25%
Encourage categorization of tags (e.g., location, subject, name)	4	17%
Offer users controlled vocabulary terms that they can re-use	2	8%

The “other” responses:

- We are currently conducting an experiment to see which works best—suggesting tags that have been added by users, suggesting tags from the bioethics thesaurus or no suggestions.
- Users create their own tags as they please.
- Users have the ability to tag content in our third party sites, e.g., blogs and Flickr group.
- None of the above though intends to show existing terms when users begin to type tag in, in the near future.
- Since the aim of the game is to score points when you match a tag of another player, tags are kept secret. In later versions, we will experiment with a list of suggested terms, based on a controlled vocabulary.
- Our API allows third-party applications to implement any model.
- Users choose their own tags. Any subject headings assigned by the library appear only in the “description” area.

Uses of Social Metadata and User-generated Content

We asked questions about the ways libraries, archives and museums intended to use the user-generated content that they were collecting. The responses are captured in the following tables (not everyone answered every question).

Table 7: Uses of social metadata (n=36)

Survey Question (Section 8)	Yes	No
Are you concerned with how the content of your site is used or repurposed?	28%	72%
Have you incorporated metadata (including tagging) created by users into your own metadata and description workflow?	39%	61%
Do you incorporate other user-contributed content (e.g., photographs, documents) into your site?	44%	56%
Does your system index user-supplied metadata?	61%	39%
Do you perform any spell-checking of user content or de-duping of tags submitted by users (e.g., differences in capitalization or spelling, singular vs. plural, etc.)?	19%	81%

Table 8: Uses of user-generated content (n=35)

Survey Question (Section 8)	Yes	No
Do you inform users in your terms of use or policy what you plan to do with the content they contribute?	77%	23%
Do you need to convert/transform user-contributed tagging/content to incorporate it into your own metadata/content?	40%	60%
Is user-supplied content searchable?	71%	29%
Do you make corrections to your existing metadata as a result of user contributions?	57%	43%
Do you have a formal process for making changes as a result of user-supplied content?	37%	63%

Most respondents (72%) were not concerned about the way the site's content is used or repurposed. We speculated that the individuals in an organization who are most concerned with data privacy and security were not those who responded to the survey. The preponderance of education-focused sites represented by the survey respondents may also be a factor in the lack of concern about how sites' content is used or repurposed, since sharing

content fits more naturally into the broader ethos of sharing knowledge and information. Sites that focused on music content were among those who expressed concern about the way the content was shared. Sites where scholars share their original work also have some concern.

More than a third (39%) incorporated user-created metadata into their own descriptive processes. We would have expected more would, since 60% of all respondents said that improving description was one of their key motivations for offering social media features (see Table 2 on p. 13).

A slightly larger minority (44%) incorporated user-contributed content such as photographs and documents into their sites. The content they incorporated fell into three categories:

- Uploading images
- Uploading, with vetting of contributions for technical reasons
- Uploading, with vetting of contributions for suitability

Respondents that use hosted services may choose not to incorporate user-contributed content into their own site but keep content at the host, such as a Flickr stream.

40% of respondents report the need to convert or transform user contributions into a format that can be used by the institution. This could speak to the varied nature of user contributions and the desire to standardize formats (for long-term preservation, for example). Similarly, the need to convert contributions might reflect the development resources available within LAMs to structure either their social metadata tools to output data in a usable format, or to adapt their repository tools to accept what the social metadata tools export.

The type of contribution also may make a difference. If the user-contributed content consists of image files, for example, these may be simply displayed. Other file formats might require different processing. Furthermore, vocabulary applied to an item might need to be standardized, even if not to the extent of developing a fully controlled vocabulary, to be effective in discovery environments.

Thirty-two respondents answered an open-ended question about how they used social networking content for description of their assets or collections. More than a third (37%) of respondents have not yet started or don't use social metadata to describe content. For the almost two thirds (63%) who do use social metadata, most seem to display them alongside their own descriptions rather than supplant local terms with those provided by users. Many also use social networking for correcting their own descriptions.

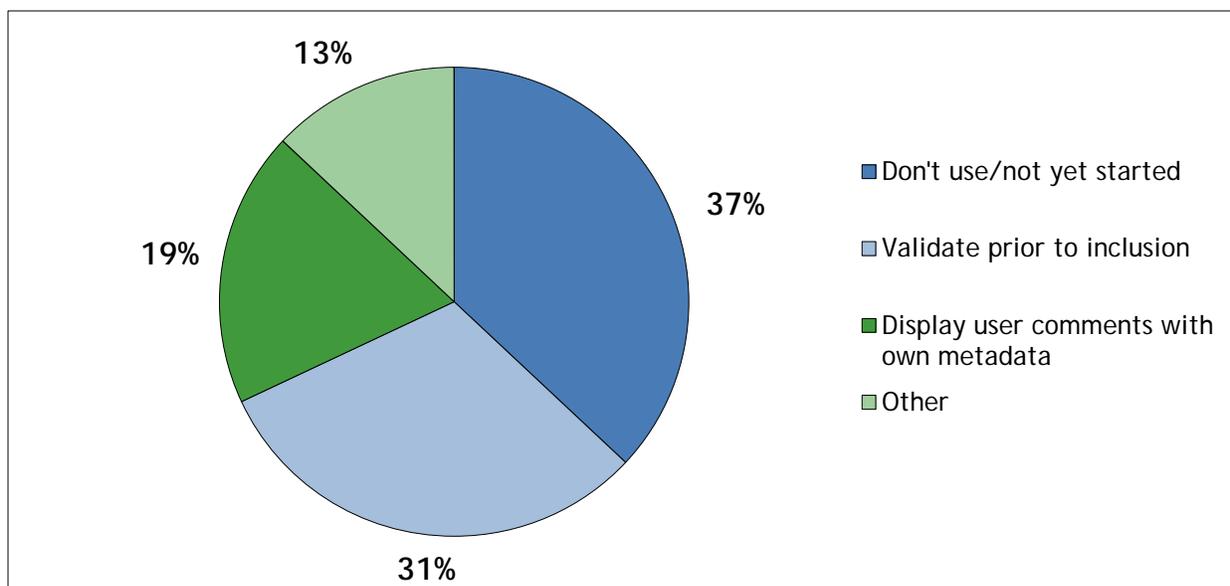


Figure 6: How respondents use social networking content for describing assets or collections (n=32)

The majority of respondents (61%) index user-supplied metadata. The determining factor seems to be the kind of user content being requested. Most of the sites that support user tagging responded that they are indexing it. Tagging is a category of social metadata that is inherently easier to index than non-textual content. An even larger majority (71%) make user-supplied content searchable. This may reflect that user-contributed content may be searchable by a web search engine rather than a formal index structure on the site.

Even though the majority of respondents are indexing user-supplied metadata and content, a larger majority (81%) of respondents perform no checks on submitted metadata such as spell-checking or de-duplication of tags. This may be because sites do not have the resources to do the moderation and validation manually or to develop a program to do so automatically. A contributing factor is whether the site suggests terms, which would obviate the need to do spell-checking. Five of the six sites that provide a list of suggested terms when users tag a resource do not do any spell-checking.

A small majority (57%) will correct the original metadata as a result of user submissions. This indicates that people are looking at user contributions and that at least some are taking user contributions seriously. Most sites (63%) indicate that they do not have a formal process for making changes to their existing metadata as a result of user-supplied content. Of the 20 sites that make corrections as a result of user submissions, 12 (60%) have a formal process for doing so. Of the eight that do not, five are relatively new, offering social media features for two years or fewer. It's possible there has not been enough time to develop formal processes around these workflows.

Tools and Functionality

Karen Smith-Yoshimura

We asked site managers whether they used open-source software, commercial software, software developed internally, or some combination. More respondents develop software internally (66%) than use open-source software (59%). Only six sites of the 29 who responded to this question (21%) use commercial software, leading us to speculate that it's not widely known what's available.

We analyzed the responses and realized that about half of the survey respondents are using both open-source software and developing software internally.

Table 9: Types of tools used (multiple responses allowed) (n=29)

Type of Tool	Number of Responses	Percentage
Developed internally	19	66%
Open-source software	17	59%
Commercial software	6	21%
Developed internally and open-source software	9	47%
Developed internally not open-source software	10	53%
Open-source software and developed internally	9	53%
Open-source software not developed internally	8	47%

We wanted to know more about the experiences of the open-source software users; why they chose it, what kinds of changes they needed to do to customize it for their needs, how much time these changes took, and whether they would recommend their open-software choices to others. We sent a follow-up questionnaire to the seventeen open-source respondents in April 2010 and received seven responses, summarized below.

Open-Source Software

Factors considered in choosing open-source software: Functionality and strong developer community support were mentioned by several. Other factors included development history and anticipated development futures; documentation and tutorials; in-house expertise; and cost. Most also cited an institutional commitment to contribute to open knowledge and open content and “stay in that ecosystem.” A key reason for using open-source software is that “it allows a wider network of experts to share their expertise with each other.” One noted that using and developing open-source software was also a requirement by their funder.

Objections: Most respondents noted that there had been no objections to their decision to use open-source software. A couple noted that there was no clear commercial alternative that fit their needs and the choice was really between open-source software and developing everything themselves. A couple noted skepticism that a “free” software package could be as functional as a commercial package, the need for ongoing development expertise and inability to outsource maintenance responsibility.

Open-source software used: Apache and MySQL were mentioned by several. Others noted: Lucene, SOLR, Jetty, Linux, Zend PHP framework, Drupal, Dojo Ajax, and Komodo as a development environment. One site uses the [Everything2 search engine](#) (Everything2).

What open-source software is used for: Content management and social media features were the prime uses reported. Lucene/SOLR is used for searching; Jetty as a web server container; MySQL for storage; Linux as an operating system. Other functions noted: metadata aggregation, search indexing, collaborative work tools and features—forums, calendar and personal information management tools.

Time spent working with open-source software: Most noted they spent the bulk of their time customizing the open-source software, including integration with existing standards, metadata structuring, and transformation. Some comments:

We used the open source software as a tool/library—we didn’t change its source code at all. Our time was spent developing the applications that used it. We did not integrate it with proprietary systems because we wanted to share it as code afterwards.

The majority of the time was spent on improving the functionality of the modules we were integrating into Drupal Core as well as developing new modules to add functionality that didn’t yet exist. Some time was spent on theming our Drupal site to meet the user interface as designed by our graphics staff. We have not yet had to spend a lot of time integrating it into other systems; however, we will be working to tie our Drupal site to a D-Space repository this summer.

Estimated time spent on developing open-source software before and after launch: This proved difficult to answer. Time spent before launch ranged from one month of full time work for one person to two full-time developers for three years. One estimated 1,000 hours of development time before launch and 250 hours after launch. Another estimated that they invest a few hours per month since launch.

Right choice? Everyone thought they had made the right choice in selecting open-source software and most would recommend their choices to others. One thought they might now choose something Ruby-on-Rails or Python-based rather than PHP and the Zend framework, although, both noted that their current systems are “working perfectly adequately.” Some comments:

When we used proprietary [software] or built it ourselves we couldn't share our systems afterwards which [we] considered a great shame....we want to share our stuff it cost us a lot to do to help others who don't have such good funding/developers.

I would recommend it to others, but they need to have the technical resource to manage and support it. Or make sure they have the budget to pay for support from external providers.

At the time we made the right choice. This is an unfair question; there are so many more content management systems available now. I would not recommend the Everything2 engine to others, but had it survived it is actually a better and more flexible content management system than Drupal but it has more overhead.

Functionality

Most sites (66%) do not generate suggestions to users about similar content on the site. Of the minority that do, six (19%) generate suggestions based on what the user is looking at; five (16%) generate suggestions based on the terms used in users' searches; three (9%) generate suggestions based on the frequency of clicks. One respondent noted that users can share bookmarks to suggest content to others with a plug-in. One motivation for user-contributed content is to help people find things that are not obviously alike. If you provide this functionality, you don't have to try to do that for them. There also needs to be a critical mass of content to make this kind of suggestion useful, and most of the sites are relatively new.

Not all the sites support searching. Of the 31 sites that do, six (19%) reported that they do not rank search results, but one noted that “weighting algorithms are pending.” Relevance ranking is by far the most common (68%); most recently added is the next most common ranking mechanism (16%). “Most commented” was used by only one site; we would think comments would indicate items that are of most interest. We speculated that people just use whatever options the system offers and “turn them on.”

Table 10: How search results are ranked (multiple responses allowed) (n =31)

Ranking Method	Number of Responses	Percentage
Relevance-ranked	21	68%
Other	13	42%
Search results are not ranked	6	19%
Most recent	5	16%
Most commented	1	3%

Those that use Flickr noted ranking by most interesting. A couple offer alphabetical listings. Other options mentioned were date of the resource and by type of content. The importance of ranking also depends on how much content is available on the site.

Reference

Everything2. 2009. "Everything Engine." <http://everything2.com/title/Everything+Engine>.

Moderation

Carol Jean Godby

Moderation consists of three broad activities: editing and approving user submissions before they are posted, deleting spam, and responding to users who abuse a site by trying to add inappropriate content.

This section of the survey started with a deceptively simple question: “Do you moderate any user contributions to the site?” 27 of 36 respondents (75%) responded yes, a higher than anticipated percentage. A closer reading of the data indicates that some respondents considered “monitoring” and “moderation” to be synonymous, which may have contributed to the high positive response rate. The working group distinguished the two: “moderating” implies user-supplied content is filtered for relevance or propriety; “monitoring” represents a broader set of activities, which might include checking the site for evidence of third-party participation and devising outreach strategies for increasing it.

There is overlap among respondents who said that they moderate user contributions, those who said they encouraged user communities, and those who were involved in “nurturing a community through outreach activities.” But only nine sites responded affirmatively in all three of these categories. We could discern no patterns in the type of sites that moderated user contributions when we looked at the resources or staff time devoted to the site and categories of staff dedicated to the site.

36% of the respondents approved user submissions before they were posted and 50% edited user submissions. These percentages seem high, since approving and editing submissions are labor-intensive. Ten of the 18 that edited user submissions also approved them before submission. We looked at the volume of user contributions reported by those who edit user contributions: 72% have fewer than 100 contributions per month, 17% have 100-499 contributors per month, and only one had more than 1,000 contributions per month. For sites where the user contribution rate is normally light, editing contributions would not be as time-consuming as we might otherwise think.

Table 11: Moderation (n=36)

Survey Question	Yes	No
Do you moderate any user contributions to the site?	75%	25%
Do you approve any user submissions before they are posted?	36%	64%
Do you edit any user contributions?	50%	50%

The management of abuse is linked to sites' requirements before users can add content. Requiring users to log in, display identifiable usernames, and submit their contributions to editorial review might minimize abusive behavior. Supporting this interpretation is an approximately 50% correlation between the requirement to login and the blocking of users. About half of all sites (53%) require a login or registration. Those that require registration generally also make the user names publicly visible. Some sites, including the Australian Newspapers project, offer searching without a login but do require a login for contributions. Many (47%) display users' names, but respondents may be thinking of user names that differ from the person's true identity, an option offered by 36% of the respondents.

Table 12: Tracking users (multiple responses allowed) (n=36)

Action	Number of Responses	Percentage
What do you track or require users to do before they can add content to your site?		
User must login/register	19	53%
Users' names are publicly visible in association with their contribution	17	47%
User must enter CAPTCHA phrase to add content	13	36%
User must provide valid e-mail address	13	36%
User name may be different from true identity to protect user's privacy	13	36%
Our site tracks the IP addresses of all contributors	11	31%
No login or registration required—users are allowed complete anonymity	11	31%

The monitoring practices are apparently successful because the spam and abuse rate is reported to be low. Only two sites reported that spam represents a serious problem; nine reported spam as an "occasional problem." Cultural heritage organizations seem to be unlikely spam targets. Only 36% of the respondents reported abusive user contributions, which happened a few times a year or less in over half of the cases. Only three sites—AcaWiki, Digital NZ Search and WorldCat.org—reported abusive contributions as often as "a few times per week." More than half of the respondents who reported abuse on their sites blocked future contributions after the first infraction. These results imply that abusive user

behavior is sporadic and easily managed, which should be especially encouraging to resource-strapped cultural heritage institutions.

Table 13: Spam (n=36)

Survey Question	Yes	No	Occasionally
Is spam a serious problem on your site?	6%	69%	25%

Table 13.5: User abuse (n=36)

Survey Question	Yes	No	Not Reviewing Content
Have users abused your site by trying to add inappropriate contributions?	36%	61%	3%

Staffing and Site Management

Helice Koffler and Karen Smith-Yoshimura

We had expected that at least a few sites would not have identified formal roles for their staff, but all sites had staff performing some role. Administration roles predominate, with monitoring the site being the most common role staff serve on respondents' sites (89%). The second most common role was answering questions and adding information and details to the site (80%). We were encouraged by the relatively high percentage that reported that their staff engaged in participatory roles, not just administration: nurturing or furthering a community through outreach activities (49%) and participating in conversations with users (46%). However, since "building a community" was a key objective for 81% of the sites (see table 2), we would have expected a higher percentage of staff serving a "nurturing a community" role.

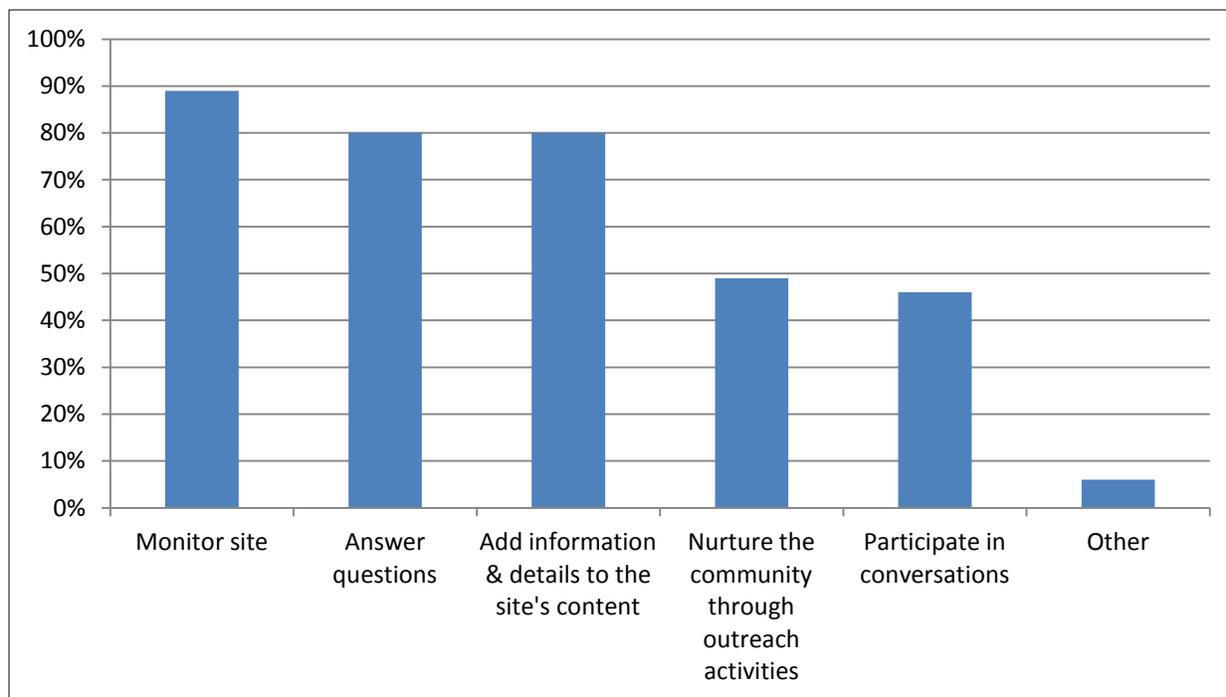


Figure 7: Roles staff serve on site (n=35)

Three of the thirty-five respondents reported that managing the site was the primary responsibility of a professional member of the staff. These three sites represented a discipline-based wiki for the scholarly community, an academic library special collections blog, and a wiki developed and maintained by the National Archives, Great Britain. We had expected that there would be a higher percentage of sites that gave managing sites to be “part of the duties” of professional staff (defined here as archivists, curators or librarians) than the 57% reported. In addition to traditional LAM professionals, 57% of sites also reported that information technology staff played a role in site management.

The use of volunteers of various types was quite high (23%), with interns comprising 14%. Museums and historical societies have an established tradition of using volunteers.

The survey also revealed there to be a considerable amount of “other” types of staff involved with social media sites. These include communications staff, customer services, assistants, exhibit development staff, museum front-line staff, and digital staff. We grouped these responses as “organization staff (non IT),” representing 34% of the responses. The variety correlates with the relative newness of these sites and implies a degree of experimentation. Roles have not yet been formalized or become mainstream.

**Table 14: Staff who play a role in site management
(multiple responses allowed) (n=35)**

Staff Type	Number of Responses	Percentage
Part of the duties of professional archivist/curator/librarian	20	57%
Information Technology Staff	20	57%
Organization staff (non-IT)	12	34%
Volunteers trained by staff	8	23%
Interns	5	14%
Primary responsibility of professional archivist/curator/librarian	3	9%

37% of sites responded that their staff spends more than 20 hours a week in activities related to creating and maintaining the site, although only 9% had staff whose primary responsibility was site management. However, this apparent disparity could be accounted for if that figure represents a combination of people, whose total hours spent on the site added up to over 20 hours a week. We were also encouraged that 20% of respondents reported that staff time dedicated to the site averaged between 6 to 19 hours per week.

We considered whether so many spending fewer than 6 hours a week (42%) may indicate the institutions’ level of commitment to the site, but decided that it is more likely an indication

that the site is still in development. We infer that more time and effort are invested in the beginning, during the launch and building of the site, than is required later, when the site is well established and less content is being added.

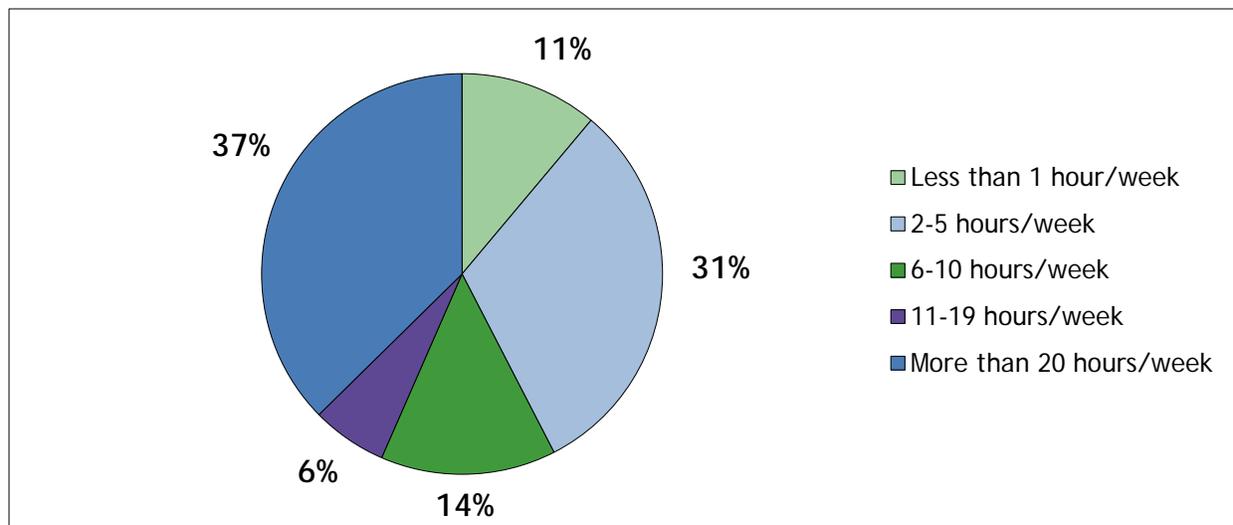


Figure 8: Staff time committed to creating and maintaining the site (n=35)

We asked respondents to estimate the percentage of time, rounded to the nearest 10%, that goes into maintaining the site. On the whole, staff seem to split up their time across several activities. Responses depended on the nature of the site, as well as the age of the site.

Table 15: Percentage of time spent on different site maintenance activities (n=30)

Activity	Percentage of Time Spent										
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Maintaining the site	3	12	7	4	1	0	0	0	2	1	0
Adding new content	1	9	4	3	3	3	2	2	0	2	0
Moderation	2	16	1	4	1	0	0	0	0	0	0
Incorporating user generated content	9	5	4	1	0	0	0	0	0	0	0
Adding new features or modifying the site's interface	5	4	6	3	2	0	0	0	0	0	0
Planning and administration	1	9	9	1	3	0	0	0	0	0	0

Additional analysis divided the sites into two categories: new (social media features less than three years old) and mature (social media features three or more years old). Sites whose social media features were “new” represented 69% of the total; sites whose social media features were “mature” represented 31%. The average time spent on each activity was calculated by category. The major disparities were in “Adding new content” (new sites=14%; mature sites=37%) and “Moderation” (new sites=6%; mature sites=12%).

Site Management

We asked whether the site was associated with a single organization or was the result of a collaboration among two or more organizations (and if the latter, to indicate how the responsibilities for managing the site were negotiated). There was a fairly even split reported between the 60% of sites emanating from a “single” institution or organization (21/35) and the 40% that were the result of a collaboration among two or more groups (14/35). At least half of those sites with multiple collaborators are structured in such a way that one organization is responsible for the administrative and maintenance functions of the site, while the other organizations involved focus on providing content.

These collaborations are often complex undertakings. For example, [DigitalNZ](#) noted in their response that their site had a “full time relationship manager.” [Waisda?](#) has four collaborators: The Netherlands Institute for Sound and Vision, Dutch Broadcaster KRO, Q42 (Internet developer) and the Free University Amsterdam. The Netherlands Institute for Sound and Vision is mostly responsible for project management and the content, the KRO for publicity and content, Q42 for the development of the website and implementing changes, and the Free University for the research. However, all partners help out with all activities.

Many sites (38%) added new content daily (see figure 9). 83% of the sites added new content on at least a monthly basis. It was clear that the sites that responded to the survey are fairly active, current sites. Only one of the sites, [Argus Index](#), a relatively early effort from the National Library of Australia that relies largely on volunteers to index a decade of articles in a defunct newspaper, is not currently adding new content.

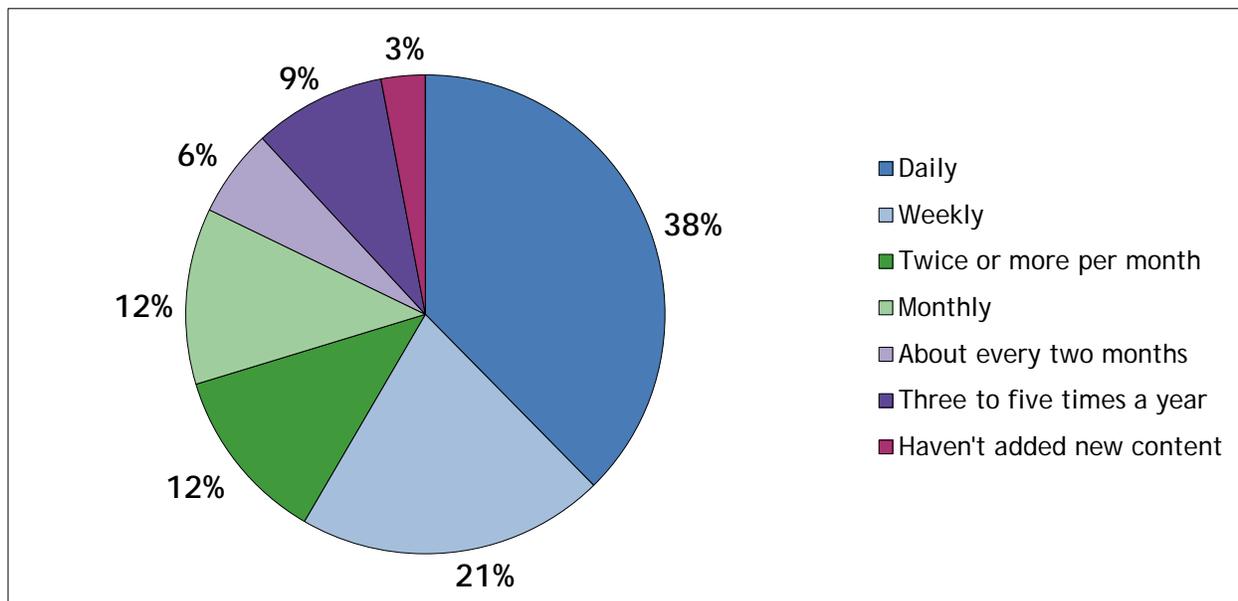


Figure 9: Frequency with which new content is added to site (n=34)

We asked site managers to comment on whether they thought their project was sustainable. A high percentage of respondents (74%) thought their projects were sustainable (see table 16). We suspect that had the survey been taken a few years ago, when there was a lot of concern about funding and far fewer social media sites existed, these managers would have been less optimistic about their sites' futures.

Eight respondents were unsure of their project's sustainability. Only one did not think the site was sustainable, noting that "we are currently looking for another platform for the site. We are at the end of our technology cycle." This site was one of the most mature of the sites that responded to the survey and that we reviewed, having offered social media features for more than four years.

Responders identified the following as contributing factors to the sustainability of their sites:

- Extensive automation of processes involved with the site's maintenance
- Reliability of host site
- Continued funding
- The need to adjust current staffing or workflows
- The infrastructure we've developed is an invaluable R&D tool, as well as a visitor engagement tool; we're very well indexed by Google.

Respondents unsure of their site's sustainability noted these factors:

- New priorities requiring reduced staffing allocated to the site
- Need for continued funding and IT support
- Dependency on the marketplace; commercial sites can fail without notice. Changes in our audience.

We found it reassuring that so many are heading towards integrating the sites into their mainstream services rather than being dependent on external or temporary funding sources.

We asked whether the organization's workflow changed since implementing the social media features on their sites and, if so, to outline the key points about how it changed. We had expected that the "yes" responses would be higher than the 60% of 35 respondents recorded. However, the majority of respondents represent newer sites; most likely not enough time elapsed for workflows to have been analyzed and changes implemented.

Where workflow adjustments had taken place, several recurring themes emerged:

- Several noted new responsibilities for staff to interact with the public and to engage in social media. This role requires new skills, some of which may be uncomfortable for existing staff.
- Adjust digitization priorities based on usage, develop more content for web.
- New responsibilities for monitoring the site, including moderating contributions.
- Redesign of cataloging procedures. For example, a subject term review process was adjusted.
- As participation grows, new tools to manage ingest and processing may need to be developed.
- New consciousness of the site as a vehicle for content, so items that had been handled traditionally for one purpose, now were being repurposed for social media sites as well.

Clearly, workflow adjustments depend on how many comments or other user generated content a site receives. If the site does not receive many contributions, the impact upon workflow would be minimal. As noted earlier, moderation and adding new content are two of the most frequent roles performed by staff, especially for mature sites.

Table 16: Site management (n=35)

Survey Question	Yes	No	Unsure
Do you think your project is sustainable?	74%	3%	23%
Has your workflow changed since implementing your site?	60%	40%	

Policies

We asked what types of policies or guidelines the site managers have implemented, offering a list of nine options and an open-text box for other (see table 17). The vast majority of respondents (77%) have an upfront policy informing users what they plan to do with the content they contribute, as might be expected from cultural heritage institutions. Only four of the 35 respondents (11%) said that they had not implemented any policies.

The majority of sites (63%) were concerned with appropriate behavior. 57% of the sites retain the right to edit or remove content; almost all of them are sites that incorporate user content. We expected that these would be key concerns among the cultural heritage organizations involved. 57% of the sites report safeguarding private information, a percentage that we would have expected to be higher.

Only 31% of sites indicated that users who violate their policy about inappropriate behavior may be blocked from site. It was not clear how many of the eleven sites that responded this way actually do so. In an earlier question, thirteen sites reported that they had experienced users trying to add inappropriate contributions, and only nine reported blocking users (five of them after just one inappropriate contribution.)

Respondents were almost evenly split between retaining ownership of user-generated content by the site/institution (11) and by the contributor (9). Few had guidelines on the nature of the contributions (5).

Table 17: Types of policies or guidelines implemented (multiple responses allowed) (n=35)

Type of Policy or Guideline	Number of Responses	Percentage
Acceptable community behavior and appropriate content	22	63%
Institution retains right to edit, repurpose, or remove user-generated content without notice	20	57%
Institution safeguards users' personal information	20	57%
Responsible re-use of site's content	11	31%
Users who violate policy may be blocked from site	11	31%
Ownership of user-generated content is retained by site/institution	11	31%
Ownership of user-generated content is retained by and attributed to contributor	9	26%
Guidelines on format, controlled vocabulary, or general nature of contributions	5	14%
None—we have not implemented any policies	4	11%
Trusted users may be whitelisted (contribute without moderation)	3	9%
Other	1	3%

The other response was a policy about discussion topics.

Only 22 of the 31 sites that had implemented policies provided a link to their policies. We wondered why so few had. Did it make a difference if the site policies were extensions of existing institutional policies or new ones created as the result of situations arising from the site? Respondents were almost evenly split: 17 had policies that were extensions of existing institutional policies and 19 had created new ones as the result of situations arising from the site. One commented, "When we realized we should have a privacy policy, we wrote one." Eight sites indicated that their current policies are a mixture of new and old; all eight provided a link to their policies.

Those that use hosted sites refer to those sites' policies, such as Flickr and YouTube. Oregon State University customized its own Flickr site's policy a bit, [No Known Copyrights Restrictions](#).

Although the policies we reviewed were all different, ranging from a few sentences to formal multi-page legal documents, they reflect the shared concerns of LAMs that are opening their content to social interaction. LAMs are making efforts to maintain a safe environment for users (with particular attention to under-age users) by encouraging and enforcing acceptable community behavior and appropriate content; safeguarding users' privacy; indemnifying or otherwise protecting the institution; and upholding professional ethics and laws, particularly in regard to providing equal access and protecting intellectual property rights.

We asked whether sites' policies had impact on staffing, workflow, or relationships with existing user communities and donors. Only 25% of the 28 who responded to this question reported that there had been some impact. A higher percentage (36%) did not yet know the impact, which may reflect the relative newness of the responding sites. Given that many of the policies sites implemented are passive (retaining rights, safeguarding personal information, content ownership), we surmise that they would not have much impact on workflow. Two respondents noted that work on site policy drove revising other policies. One of these sites even had added a new communications staff member as a result of the site's activity.

Examples of Different Policies

The **State Library of Queensland** developed a new [User Online Comments Policy](#) that covers purpose, scope, general principles, examples of inappropriate comments, breaches of policy, filtering and the State Library's rights.

DigitalNZ developed new [Terms of Use](#) covering the users' responsibilities, tools, acceptable use, trademarks, content disclaimer, external links from its website, Crown copyright material, third-party copyright material, user contributed metadata, linking to its website, privacy, tracking of website usage, security and damage, disclaimer, cookies, suspension of access, and governing law. It notes that users grant the National Library a "worldwide, non-exclusive, royalty-free, perpetual, sublicenseable and transferable license to use, reproduce, distribute, modify, and display your User Contributed Metadata. You understand and agree that this can be redistributed in any format and through any channel, including the DigitalNZ websites and the DigitalNZ API." The terms of use also has links to specific terms of use for its developer's API, content contribution, and shared repository.

University of California, San Francisco's Legacy Tobacco Documents Library developed three new policy statements:

- A brief [legal document](#) noting that the website may contain copyrighted material and that each user is responsible for complying with applicable copyright laws. "By accessing this website, the user agrees to defend, indemnify and hold harmless the University of California, its affiliates and their directors, officers, employees and agents from and against all claims and expenses, including attorneys' fees, arising out of the use of this website by the user."
- A [document](#) describing the origins, format, and policy of the data, noting that the library will not alter any document images except to "redact" Social Security numbers or personal bank account numbers flagged by a notice when redaction occurs.

- A brief [Privacy Information](#) document noting the information that is automatically collected and stored such as Internet domain, type of browser, pages viewed and that the library does not track the searches or document requests of individual users.

Science Buzz has four sets of policies and guidelines that are extensions to existing ones and new ones:

- [Basic requirements for contribution](#), with a link to the “Community Guidelines Style Guide.”
- [Community Guidelines](#), that outlines the few circumstances under which they will not allow entries: offensive or inappropriate comments/stories, definition of what’s considered offensive, not related to current science or a blog entry or the website itself, spamming. It also requests attributing the source when quoting information and link to the original content and states that personal information will not be given out without permission. It links to a separate statement to the Science Museum’s statement on the scientific theory of evolution, noting that it reserves the right to “delete comments that misrepresent the principles of science.”
- [Copyright information](#), covering use of material on the Science Buzz site, sites linked from the website, and limitation of liability.
- A longer [Privacy Statement](#), covering notification of policy changes, information collected by the museum, information submitted by third parties, personal information collected by the museum online, other information collected by the museum online, cookies, SSL encryption, children’s privacy on the Web, use of information, sharing information, use of postal address information and opt-out, use of telephone numbers and opt-out, e-mail correspondence, newsletters, and informational updates, reviewing information, third-party mailing lists, and compliance.

[Policies that Extend Existing Institutional Policies](#)

The PowerHouse Museum extended its existing policies to cover [rights and permissions](#) for its content, including works under copyright restrictions, no known copyright, public domain, and creative commons.

MTagger also extended existing policies for its [Guidelines for Use](#), a brief statement on privacy (user names and web pages people tag are visible to all but people can hide user names and choose to tag anonymously) and appropriate use. MTagger users agree to the campus-wide [Information Technology Policies of the University of Michigan](#), which links to

more detailed documents describing proper use, privacy and security, data management, digital copyright, information technology system standards, student information technology policies, and related University websites.

The British Library extended existing policies for its **Archival Sound Recordings** site. Its [Terms and Conditions](#) covers posting of material, rights and permissions, warranty and indemnification, and applicable law. There's a link to a separate set of [Rules](#), such as materials must be in English; a list of restrictions on what materials may not contain such as e-mail address or contact information, offensive, misleading, or irrelevant content, advertising; contributors cannot impersonate another person; materials must not appear to replicate existing material. Like MTagger, the Archival and Sound Recordings "Terms and Conditions" links to the broader [Terms of Use](#) of the British Library covering copyright, acceptable use policy, privacy (web log files, data protection, other websites, Google analytics), legal disclaimer, links, and notice and takedown policy.

Examples of New Policies

Minnesota Reflections has a brief [policy](#) that it "improvised:"

This part of Reflections has been set up to allow the public to comment on items in the collections. You can leave whatever comment you like, though we do ask that comments be appropriate to the topic at hand.

We do monitor comments and reserve the right to change them or remove them altogether if they are offensive or inappropriate. In particular, please avoid profanity, pornography, and hate speech.

The Australian Music Centre has two new online policies:

- [Privacy Policy](#), covering collection of information, use of anonymous information, use of personal information, and access and correction to personal information.
- [Terms and Conditions](#), covering disclaimer, copyright, external links, AMC members and registered users, indemnity, submissions of material and other information, and refunds.

Open Context's [Data Publication Guidelines for Contributors](#) includes recommendations for data preparation, data formats and structures, location information and site security, copyright and licensing. It also has a new four-page [Open Context Privacy Policy](#), dated 21 April, 2010, which it developed to be "consistent with the American Library Association's guidance for protecting patron rights for privacy, confidentiality, and academic freedom." The policy's introduction:

Our commitment to your privacy and confidentiality has deep roots not only in law but also in the ethics and practices of scholarly dissemination. Open Context follows the American Library Association's Code of Ethics by protecting each user's "right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired, or transmitted.

Policies that Combine Extensions to Existing Policy and New Policy

[EthicShare's privacy policy](#) combines extensions to existing policy and new policy. It adheres to the University of Minnesota's online privacy policy, since the site is hosted and supported by the University, but amended it for EthicShare where appropriate. It covers network traffic logs, web visit logs, cookies, information voluntarily provided by users, third parties, security measures, and opt-out.

Your Archives in the UK has three sets of policies, a combination of extensions to existing policies and new ones:

- [Terms and conditions of use](#) covering scope, purpose, reuse of content, disclaimer, limitation of liability, contributions, and rules.
- [Use and re-use of images](#), covering submitting images, reuse of images, and that the public records in the National Archives are in Crown Copyright.
- [Crown Copyright](#)

AcaWiki is another of the eight sites that have a mixture of extensions of existing policies and new ones. It provided five links to policies:

- [The Creative Commons Attribution license](#)
- AcaWiki's [Terms of Use](#) covering international users, minors, purpose, lack of peer review, contributions, use of content posted to the AcaWiki website, user conduct, termination, disclaimers, limitation of liability, indemnification, copyright complaints, trademarks, privacy, third party websites and content, other terms of use, and changes to the terms of use.
- AcaWiki's [Privacy Policy](#) covering principles, personal information, what AcaWiki does with personal information—user profile, postings, e-mails and newsletters, non-disclosure of personal information, security, disclaimer, lack of linking, no selling or sharing, reorganization or spin-offs, and children; third-party sites, international users, technical information, cookies, page history, user contribution, reading projects, editing summaries, discussions, e-mail, mailing lists, and changes.

- AcaWiki's policy on [Copyrights](#) which links to the Creative Commons Attribution license above.
- A [template](#) for adding a summary with guidelines on how to fill it in and a link to posting guidelines.

The **Folger Shakespeare Library**, another of the eight sites that have a mixture of extensions of existing policies and new ones, has four online policies:

- [Copyright and permissions](#), a short statement on the restrictions on downloading text or images for public or commercial use without written permission.
- [Guidelines for linking](#), a brief two-sentence statement that both encourages links for educational purposes and cultural programs and warns that it will take action to disassociate itself from links or implied relationships not in its best interest.
- [Privacy Policy](#), covering change notifications, security and credit card information, cookies, links, liability and warranty, indemnification, and opt-out procedures.
- [Notice to Parents](#), which includes its Children's Privacy Notice, reassuring parents that the Folger is in full compliance with the Children's Online Privacy Protection Act and that it has a mechanism by which "parents or guardians can, upon verification, review their child's personal information, ask to have that information deleted, and refuse to allow any further collection or use of their child's personal information."

WorldCat.org, another site with a mixture of extensions to existing policies and new ones, has two online policies:

- [OCLC WorldCat.org Services Terms and Conditions](#) covers use of data, third-party data, content, other services, trademarks, material breach, disclaimers and limitations of liability, and an exhibit applicable to third-party databases.
- [The OCLC Privacy Policy](#) covers how and what information is collected, why information is collected, third parties, opt-out features, security, access to information, and contact and notice.

References

Australian Music Centre. "Privacy Policy." <http://www.australianmusiccentre.com.au/about/privacy>.

———. "Terms and Conditions." <http://www.australianmusiccentre.com.au/about/terms>.

British Library. "Terms of Use." <http://www.bl.uk/aboutus/terms/>.

———. "Rules." *Archival Sound Recordings*. <http://sounds.bl.uk/TextPage.aspx?page=rules>.

———. "Terms and Conditions." *Archival Sound Recordings*. <http://sounds.bl.uk/TextPage.aspx?page=terms>.

EthicShare. "EthicShare Privacy Policy." <https://www.ethicshare.org/about/privacy>.

Minnesota Digital Library. *Minnesota Reflections*. <http://views.mndigital.org/>.

National Library of Australia. 2005. Argus Index. <http://www.nla.gov.au/argus/>.

National Library of New Zealand. *DigitalNZ*. <http://www.digitalnz.org/>. Accessed 9 March 2011.

———. "Terms of Use." Accessed 10 March 2011. <http://www.digitalnz.org/terms>.

Open Context. "Data Publication Guidelines for Contributors." <http://opencontext.org/about/publishing>.

———. 2010. "Open Context Privacy Policy." http://opencontext.org/docs/OpenContext_Privacy_Policy.pdf.

Oregon State University. "No Known Copyright Restrictions." OSU Libraries—University Archives. [<http://archives.library.oregonstate.edu/no-known-copyright-restrictions>.]

Powerhouse Museum. "Rights and Permissions." *Photo of the Day*. <http://www.powerhousemuseum.com/imageservices/index.php/rights-and-permissions/>.

Science Museum of Minnesota. "Community Guidelines." *Science Buzz*. <http://www.sciencebuzz.org/community/guidelines>.

———. "Contribute." *Science Buzz*. <http://www.sciencebuzz.org/community/contribute>.

———. "Copyright Information." *Science Buzz*. <http://www.sciencebuzz.org/about/copyright>.

———. "Privacy Statement." *Science Buzz*. <http://www.sciencebuzz.org/about/privacy>.

State Library of Queensland. 2009. "User Online Comments Policy." *About us: Policies, standards, protocols*. Policy Ref: RD-CSC-1207-V01. Last updated 30 June 2009. http://slq.qld.gov.au/about/pol/user_online_comments_policy.

University of California, San Francisco. "About the Data." *Legacy Tobacco Documents Library*. http://legacy.library.ucsf.edu/about/about_data.jsp.

———. "Legal Information." *Legacy Tobacco Documents Library*. <http://legacy.library.ucsf.edu/legal.jsp>.

———. "Privacy Information." *Legacy Tobacco Documents Library*. <http://legacy.library.ucsf.edu/privacy.jsp>.

University of Michigan. "Guidelines for Use." *MTagger*. <http://www.lib.umich.edu/mtagger/tags/welcome>.

———. "Information Technology Policies at the University of Michigan." *Information and Technology Services General Computing*. <http://www.itd.umich.edu/itpolicies/>.

Waisda? <http://waisda.nl/>.

Measuring Success

Elizabeth Yakel

We were very interested in understanding how respondents from different social metadata sites perceive their success, what indicators of success they use, what types of activities they measure, and what data are collected. Sections 12 and 13 of the survey covered these issues. Even though these sections were at the end of the survey, we still received over 30 responses to most of the questions.

Success is a difficult concept to measure. Given the variety of the sites, we asked each one to assess whether they were successful and what factors they thought were important to success. The vast majority of respondents (30 or 91%) consider their sites to be successful. This is consistent regardless of the type of institution (library, archives, museum), whether the site is managed locally or uses a hosted service, or the amount of interaction on the site.

We probed why respondents considered their sites successful or unsuccessful in two ways, first by an open ended question and then by using a multiple choice question. Nineteen of the 24 respondents to the open ended question listed user or audience factors as their biggest indicator of success as reflected in the following comments:

Although we would have liked to receive more contributions, the quality of what we were able to post has been strong. If we were to do it again, we would have opened up different ways for people to participate (comments, etc.)

Just getting our material available to a wider audience makes it successful—the social media aspects are actually not the major impetus.

Only one respondent mentioned capturing information to incorporate into the official metadata:

On one hand, we have engaged with a new audience and reached a diverse population of new users. On the other hand, our sites have given us a chance to interact with our current users and other library staff in new ways. We have gotten a few extremely

valuable pieces of information about some of our collections and have had an increase in attention directed at the Archives and diversity of our collections.

We also presented respondents with a multiple choice question to help them identify success criteria. Respondents could select all the answers that applied among the following choices: Adding new content, Engaging new audiences, Engaging audiences you did not expect, Engaging and supporting existing audiences, and Gathering additional metadata about existing content.

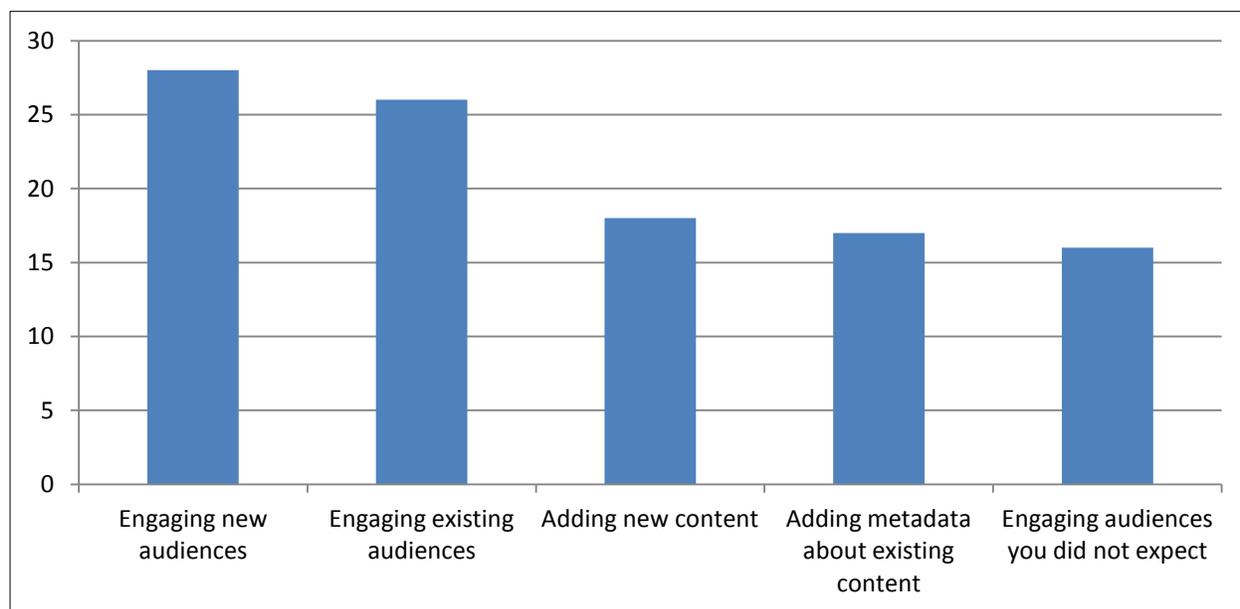


Figure 10: Success criteria (n=32)

As in the open ended questions, respondents selected both engage new audiences (88%) and existing audiences (81%) more frequently than generating new content (56%), gathering metadata (53%), and engaging unexpected audiences (50%). Respondents provided few, if any, new categories by which to judge success. A respondent from a science museum noted that “supporting exhibits with new information and experiences” was a success criteria (although this fit into what we perceived to be new content). An institution using Twitter had these goals: “Increase traffic to our [newsletter] site; increase referrals to followed accounts.”

Most respondents (95%) measure site activity and user interactions with the site. What is captured depends on the specifics of the site. The top three data elements captured are comments (76%), unique visitors (67%), and visits (64%). User names, IP addresses, uploads, and downloads are captured the least. In addition to site functionality, the collection of user names and IP addresses raise privacy issues in which some institutions might not want to be involved.

Table 18: Types of activities measured (multiple responses allowed) (n=33)

Type of Activity	Number of Responses	Percentage
Number of comments	25	76%
Number of unique visitors	22	67%
Number of visits	21	64%
Number of user accounts	19	58%
Most frequently accessed content	19	58%
Geography of users	15	46%
Number of tags	14	42%
Domain IP	8	24%
Number of uploads	7	21%
Other	6	18%
Number of downloads	4	12%
User names	1	3%
Number of groups formed	0	0%

Data capture is done using a variety of obtrusive and unobtrusive methods. Overwhelmingly, the technique of choice is web analytics. 31 of the 33 respondents (94%) use some analytics tool. All other methods of data capture are used less frequently. The other unobtrusive or naturally occurring methods of generating data (registration and user content) were the next most popular. Surveys and focus groups, which are both more obtrusive and require additional effort, are the least frequently employed techniques for generating user information.

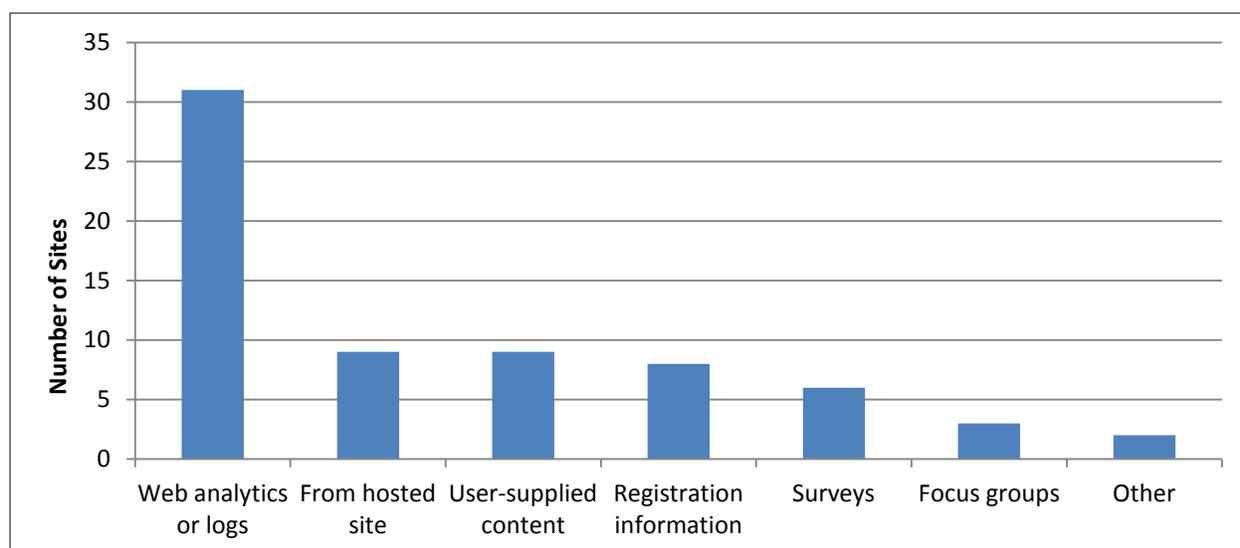


Figure 11: Data capture methods (n=33)

We also identified two potential, albeit problematic measures of success: the number of site visitors and the number of content contributors to the site per month. The total number of visitors ranged from under 250 per month to over 100,000. Half of the respondents reported 4,999 or fewer visitors per month; 21% had fewer than 1,000 visitors. On the other hand, 21% had more than 5,000 visitors per month; four sites had over 100,000 visitors per month.

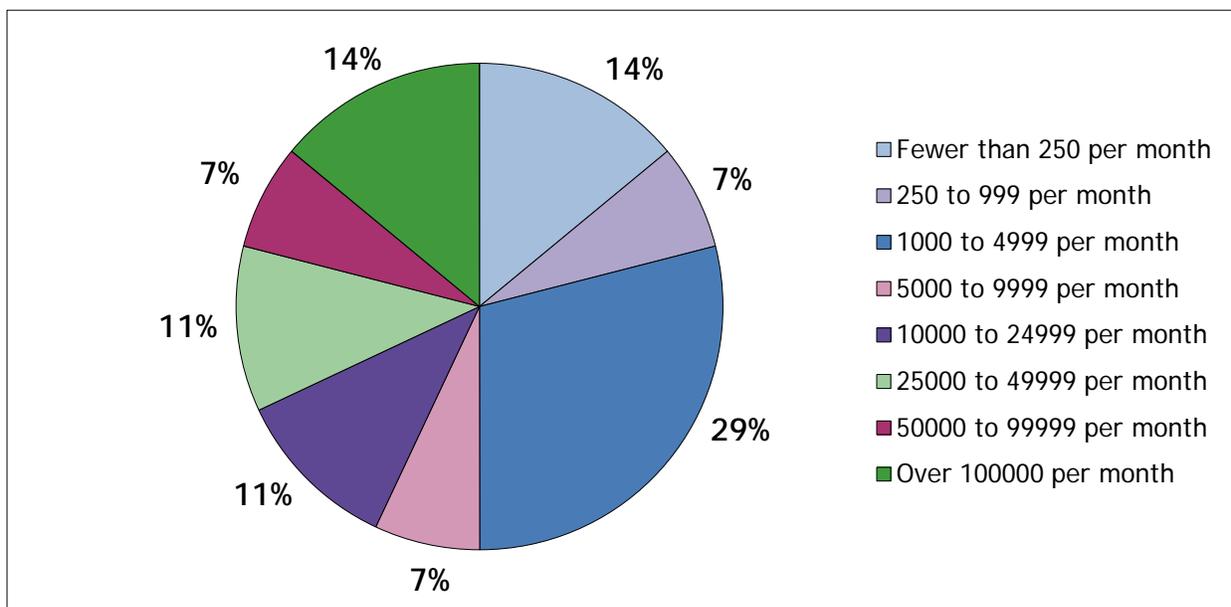


Figure 12: Frequency of unique visitors per month (n=28)

If site visitors indicate the size of the audience, content contributors represent the core of Web 2.0 activities. 20 respondents (67%) reported 100 or fewer visitors contributing content each month; 26 (87%) stated they had 499 visitors or fewer contributing content per month. Only three sites reported over 1,000 visitors contributing content—Australian Newspapers, Distributed Proofreaders, and WorldCat.org.

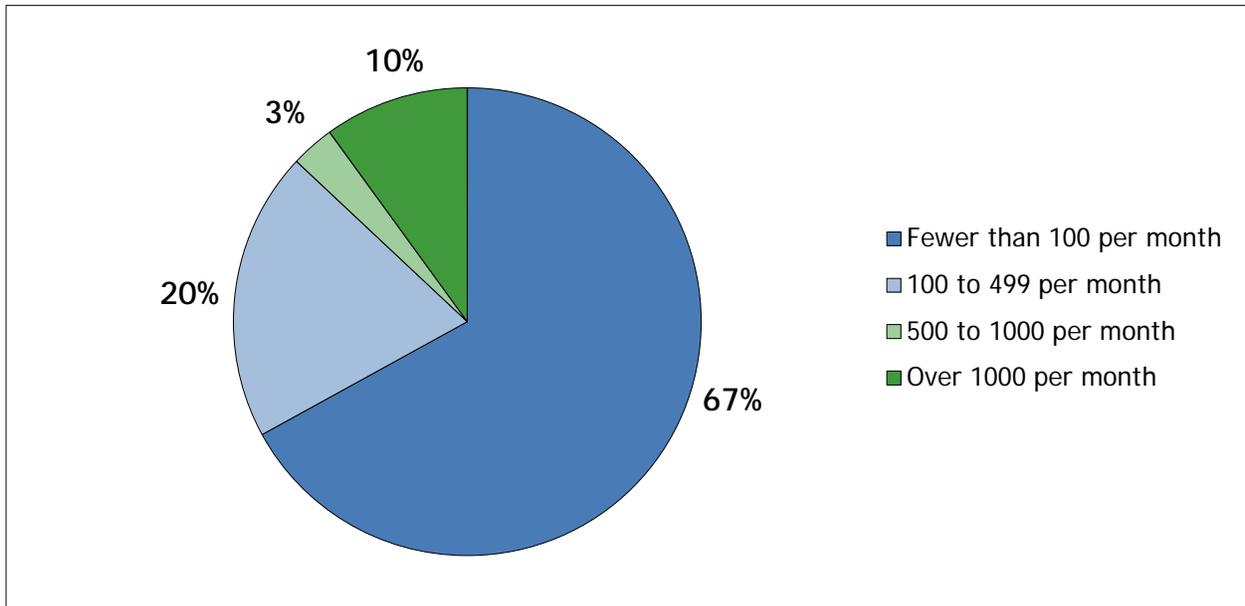


Figure 13: Visitors contributing content per month (n=30)

Neither the number of overall visitors nor the number of visitors contributing content had any relationship to whether a site perceived it was successful. Two of the respondents who identified their site as unsuccessful had fewer than 250 visits per month, but one had over 100,000; likewise two of them had fewer than 100 visitors contributing content per month but one had over 1,000.

Respondents appear to be monitoring the users of their social media sites. Use of a web analytics tool was almost universal and sites are monitoring how visitors are using the Web 2.0 functionalities they have implemented. There is only limited use of surveys and focus groups.

Engaging audiences is a major factor that respondents are using to judge the success of their sites; however, this survey indicates that engagement is being measured by quality, not quantity.

Appendix A: Sites that Responded to Social Metadata Survey at a Glance²

Responding Site	URL	Ctry	Site Type	Site Management	Types of Content on Site				
					Text	Still Images	Audio	Moving Images	Other
AcaWiki	acawiki.org	US	Discipline	Manage own	Yes	Yes			
Anteater Antics	ucisca.wordpress.com	US	Archive/Library	Manage own	Yes	Yes	Yes	Yes	
Archival Sound Recordings	sounds.bl.uk	UK	Archive	Manage own	Yes	Yes	Yes		
Argus Index*	nla.gov.au/argus	AU	Library	Hosted service	Yes				Includes internal links
Australian Music Centre	australianmusiccentre.com.au	AU	Discipline	Manage own	Yes	Yes	Yes	Yes	
Australian Newspapers	newspapers.nla.gov.au	AU	Library	Manage own	Yes	Yes			
Beyond Brown Paper	beyondbrownpaper.plymouth.edu	US	Archive/Library	Manage own		Yes			
Brooklyn Revealed	brooklynrevealed.com	US	Archive	Hosted service	Yes	Yes		Yes	
California Digital Library*	cdlib.org	US	Archive/Library	Manage own	Yes				
California Digital Library's Twitter Presence	twitter.com/CalDigLib	US	Library	Hosted service	Yes	Yes			Links to various content forms
Claremont Colleges Digital Library	ccdlib.libraries.claremont.edu	US	Archive/Library	Manage own	Yes	Yes		Yes	
Click! Photography changes everything	click.si.edu	US	Archive/Library /Music	Manage own	Yes	Yes			
DigitalNZ Search	search.digitalnz.org	NZ	Archive/Library Music/Community	Manage own	Yes	Yes	Yes	Yes	
Distributed Proofreaders*	pgdp.net	US	Community	Manage own	Yes	Yes			

² Asterisk indicates those sites not covered by site reviews in report 1

Responding Site	URL	Ctry	Site Type	Site Management	Types of Content on Site				
					Text	Still Images	Audio	Moving Images	Other
EthicShare	ethicshare.org	US	Discipline	Manage own	Yes				Aggregation of bibliographic references from disparate sources about bioethics
Find	find.natlib.govt.nz	NZ	Library	Manage own	Yes	Yes			
Folger Shakespeare Library	folger.edu	US	Archive/Library /Music	Manage own	Yes	Yes	Yes	Yes	
Fundacion Ignacio Larramendi*	larramendi.es	ES	Archive/Library	Manage own	Yes	Yes			
Galaxy Zoo	galaxyzoo.org	UK	Discipline	Manage own	Yes	Yes			
Getty Research Institute, Flickr Commons	flickr.com/photos/gettyresearchinstitute	US	Library/Community	Hosted service		Yes			
Kew Gardens	kew.org	UK	Music/Community	Manage own	Yes	Yes	Yes	Yes	Interactive maps; calendar; live content from expeditions via Twitter
Library of Congress, Flickr Commons	flickr.com/photos/library_of_congress	US	Library/Community	Hosted service		Yes			
Minnesota Historical Society	mnhs.org	US	Archive	Manage own	Yes	Yes	Yes		
MTagger	lib.umich.edu/mtagger/	US	Library	Manage own	Yes	Yes			
Minnesota Reflections	reflections.mndigital.org	US	Archive/Library	Manage own		Yes			
New York Public Library*	nypl.org	US	Archive/Library	Manage own	Yes	Yes	Yes	Yes	
New York Public Library, Flickr Commons	flickr.com/photos/nypl	US	Library/Community	Hosted service		Yes			
NSW Reference and Information Services Group*	nsw-risq.org	AU	Library	Hosted service	Yes	Yes			
Open Context	opencontext.org	US	Discipline	Manage own	Yes	Yes			Tabular data
Oregon State University's Commons and Archives, Flickr Commons	flickr.com/photos/osucommons/ and flickr.com/photos/osuarchives/	US	Archive/Library Community	Hosted service	Yes	Yes			
Picture Australia	pictureaustralia.com.au	AU	Library/Community	Manage own		Yes			
Polar Bear Expedition Digital Collections	polarbears.si.umich.edu	US	Archive/Community	Manage own	Yes	Yes	Yes	Yes	
Powerhouse Museum	powerhousemuseum.com	AU	Music	Manage own	Yes	Yes	Yes	Yes	

Responding Site	URL	Ctry	Site Type	Site Management	Types of Content on Site				
					Text	Still Images	Audio	Moving Images	Other
Science Buzz	sciencebuzz.org	US	Music	Manage own and hosted service	Yes	Yes	Yes	Yes	Flash games and other interactive
State Library of Queensland	slq.qld.gov.au	AU	Library	Manage own	Yes	Yes	Yes	Yes	Virtual page-turning books
Steve: The Museum Social Tagging Project	[Editor's Note: Unsafe URL Removed]	US	Music	Manage own	Yes	Yes			
UCSF Tobacco Industry Videos	archive.org/details/tobaccoarchives	US	Archive	Hosted service			Yes	Yes	
Waisda?	waisda.nl	NL	Archive	Manage own				Yes	
Wallerawang Branch Library	wallerawanglibrary.blogspot.com/	AU	Library	Manage own	Yes	Yes	Yes	Yes	
Wikipedia (Germany)	de.wikipedia.org	DE	Community	Manage own	Yes	Yes	Yes	Yes	
WorldCat.org	worldcat.org	US	Archive/Library	Manage own	Yes				
Your Archives	yourarchives.nationalarchives.gov.uk/	UK	Archive	Manage own	Yes	Yes			
Total (of 42 sites reviewed)					34	36	15	16	

Appendix B: Social Media and User Contributions Offered at a Glance³

Responding site	Social Media and User Contribution Features Offered																		
	Notes	Bookmarks	Collaborative Filtering	Comments	Edit Text	Form sub-groups	Ratings	Reviews	RSS	Synchronous Chat	Tagging	Upload Materials	User Awareness (who's logged on)	User Profiles	User Recommendations	User-compiled Lists	User-contributed Images	User-contributed Video	Other
AcaWiki	Yes			Yes	Yes			Yes			Yes	Yes		Yes			Yes		
Anteater Antics				Yes				Yes			Yes								
Archival Sound Recordings	Yes	Yes									Yes								
Argus Index*				Yes															
Australian Music Centre				Yes				Yes	Yes										
Australian Newspapers	Yes			Yes	Yes						Yes								
Beyond Brown Paper	Yes	Yes		Yes					Yes										
Brooklyn Revealed				Yes				Yes					Yes		Yes				
California Digital Library*	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
California Digital Library's Twitter Presence				Yes															Tweets, re-tweets

³ Asterisk indicates those sites not covered by site reviews in report 1

Responding site	Social Media and User Contribution Features Offered																		
	Notes	Bookmarks	Collaborative Filtering	Comments	Edit Text	Form sub-groups	Ratings	Reviews	RSS	Synchronous Chat	Tagging	Upload Materials	User Awareness (who's logged on)	User Profiles	User Recommendations	User-compiled Lists	User-contributed Images	User-contributed Video	Other
Claremont Colleges Digital Library				Yes					Yes			Yes							
Click! Photography changes everything									Yes			Yes					Yes		
DigitalNZ Search				Yes			Yes		Yes		Yes								Suggest additions to all metadata fields in search index
Distributed Proofreaders*				Yes		Yes							Yes						Forums are primary form of communication
EthicShare				Yes		Yes					Yes			Yes	Yes	Yes			
Find				Yes			Yes				Yes								
Folger Shakespeare Library	Yes	Yes		Yes		Yes			Yes		Yes	Yes	Yes	Yes		Yes	Yes	Yes	
Fundacion Ignacio Larramendi*	Yes	Yes		Yes			Yes	Yes	Yes		Yes			Yes					
Galaxy Zoo	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Getty Research Institute, Flickr Commons				Yes		Yes					Yes					Yes			
Kew Gardens				Yes			Yes		Yes			Yes			Yes		Yes		
Library of Congress, Flickr Commons	Yes	Yes		Yes					Yes		Yes	Yes							
Minnesota Historical Society	Yes			Yes															
MTagger	Yes								Yes		Yes			Yes		Yes			
Minnesota Reflections				Yes															
New York Public Library*				Yes					Yes									Yes	
New York Public Library, Flickr Commons	Yes			Yes					Yes		Yes								

Responding site	Social Media and User Contribution Features Offered																		
	Notes	Bookmarks	Collaborative Filtering	Comments	Edit Text	Form sub-groups	Ratings	Reviews	RSS	Synchronous Chat	Tagging	Upload Materials	User Awareness (who's logged on)	User Profiles	User Recommendations	User-compiled Lists	User-contributed Images	User-contributed Video	Other
NSW Reference and Information Services Group*	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
Open Context				Yes					Yes		Yes	Yes					Yes		
Oregon State University's Commons and Archives, Flickr Commons	Yes			Yes					Yes		Yes								
Picture Australia											Yes	Yes					Yes		
Polar Bear Expedition Digital Collections		Yes	Yes	Yes									Yes	Yes					
Powerhouse Museum				Yes					Yes		Yes				Yes		Yes		
Science Buzz				Yes	Yes	Yes	Yes		Yes		Yes	Yes	Yes	Yes			Yes	Yes	Registered users can create blog entries
State Library of Queensland	Yes	Yes		Yes			Yes	Yes	Yes	Yes	Yes	Yes				Yes			
Steve: The Museum Social Tagging Project											Yes			Yes					
UCSF Tobacco Industry Videos		Yes		Yes			Yes	Yes	Yes		Yes				Yes				Forum posts
Waisda?											Yes			Yes					
Wallerawang Branch Library				Yes					Yes		Yes								
Wikipedia (Germany)				Yes	Yes				Yes		Yes	Yes		Yes		Yes	Yes	Yes	
WorldCat.org	Yes						Yes	Yes	Yes		Yes			Yes		Yes			
Your Archives	Yes			Yes	Yes						Yes	Yes				Yes	Yes		
Total (of 42 sites reviewed)	14	8	1	32	5	5	8	8	21	1	26	12	5	11	5	8	10	4	

Appendix C: Social Metadata Survey Questions

Page 1. Site Background

This survey is targeted to social metadata/media site managers in cultural heritage institutions in North America, Europe, Australia, and New Zealand. The [RLG Partners Social Metadata Working Group](#) designed the survey as part of its effort to analyze issues around sharing and aggregating user contributions.

Your experiences are very important to our research. From your experiences we hope that the RLG Partnership and others may learn how libraries, archives, and museums could benefit more from contributions by the audiences they seek to reach.

Your site will be identified as participating in the survey and information you provide will be included in the RLG Partners Social Metadata Working Group report. Your name, however, will not be published.

We estimate that this survey will take 20 to 30 minutes of your time. Please consult with your colleagues as needed to answer questions. **Please respond by November 13.** You can return to the survey at any time before the survey is closed. For your convenience, a PDF version of this survey is available [click here](#).

Required questions are marked by an asterisk (*). There are no word limits in the free-text comment boxes. If you have any questions about the survey, please contact Karen Smith-Yoshimura at smithyok@oclc.org.

Thanks in advance! We appreciate and value your input.

1. What is the name of your site?

2. What is the URL of your site?

3. What is/are the key objective(s) for offering social media on your site? [Check all that apply]

- Increase traffic and access to our content
- Build user community
- Enhance description
- Build collection
- Other (please specify)

4. What types of content are on your site? [Check all that apply]

- Audio
- Moving images
- Still images
- Text
- Other (please specify)

5. How long have you offered social media features on your site?

- Not yet public
- Less than 1 year
- 1-2 years
- 3-4 years
- More than 4 years

6. Do you manage your own site or use a hosted service?

- Manage own site
- Use hosted service

Page 2. Manage own site—tools

1. What types of tools are you using for your social metadata site? (Check all that apply)

- Open-source software
- Commercial software
- Developed internally

Page 3. Content

1. What type of social media/user contribution features do you offer on your site? [Check all that apply]

- Annotations or Notes
- Bookmarks
- Collaborative filtering
- Comments
- Edit text
- Form sub-groups/community
- Ratings
- Reviews
- RSS
- Synchronous chat
- Tagging
- Upload materials
- User awareness (who is logged on)
- User profiles
- User recommendations
- User-compiled lists
- User-contributed images
- User-contributed video
- Other (please specify; you are not limited by the size of the box)

2. Why did you choose the social features you use in your site? (You are not limited by the size of the box.)

3. What are the sources of content on your site? [Check all that apply]

- Archives
- Commercial vendors
- Government agencies

- Libraries
- Museums
- Scholars and scholarly societies
- Users
- Other (please specify)

4. If you use a controlled vocabulary to describe your content, which one(s) do you use? [Check all that apply]

- Library of Congress Subject Headings
- Medical Subject Headings
- Art and Architecture Thesaurus
- Local
- ACRL Rare Books and Manuscripts Section thesauri
- Other (please specify)

Page 4. Tagging

1. Which uses of tagging apply to your site? [Check all that apply]

- Provide a list of suggested terms when users tag a resource
- Offer users controlled vocabulary terms that they can re-use
- Show users existing tags in the system
- Encourage categorization of tags (e.g. location, subject, name)
- Combine user-contributed tags with our own controlled terms
- Other (please specify)

Page 5. Audience / Users

1. What communities are you engaging with your site? [Check all that apply]

- General public
- Hobbyists
- Students: between ages 6 and 13

- Students: between ages 14 and 18
- Students: college and university undergraduate/graduate students
- Teachers
- Academic faculty
- Independent scholars
- Professionals
- Other (please specify)

2. Do you feel that you have encouraged user communities by allowing user-contributed content?

- Do you feel that you have encouraged user communities by allowing user-contributed content? Yes
- No

3. What evaluations have you done with site participants? [Check all that apply]

- Feedback/comment/suggestion box
- Surveys
- Focus Groups
- Interviews
- Usability testing
- None
- Other (please specify)

Page 6. Moderation

1. Do you moderate any user contributions to the site?

- Yes
- No

2. Do you approve any user submissions before they are posted?

- Yes
- No

3. Do you edit any user contributions?

Yes

No

4. What do you track or require users to do before they can add content to your site? [Check all that apply]

No login or registration required - users are allowed complete anonymity

Our site tracks the IP addresses of all contributors

User must enter CAPTCHA phrase to add content

User must login/register

User must provide valid email address

User name may be different from true identity to protect user's privacy

Users' names are publicly visible in association with their contribution

5. Is spam a serious problem on your site?

Yes

No

Occasional problem

6. Have users abused your site by trying to add inappropriate contributions?

Yes

No

Not reviewing content

7. How many inappropriate contributions do you allow before blocking that user from your site?

We don't block any users

One

Two

Three

Four or more

Page 7. Moderation - User abuse

1. About how often do users abuse your site by trying to add inappropriate content?

A few times a year or less

A few times a month

A few times a week

Other (please specify)

Page 8. Use of Social Metadata

1. Are you concerned with how the content of your site is used or repurposed?

Yes

No

2. Have you incorporated metadata (including tagging) created by users into your own metadata and description workflow?

Yes

No

3. Do you incorporate other user-contributed content (e.g., photographs, documents) into your site?

No

Yes (please indicate how you do that; you are not limited by the size of the box)

4. Do you inform users in your terms of use or policy what you plan to do with the content they contribute?

Yes

No

5. Do you need to convert/transform user-contributed tagging/content to incorporate it into your own metadata/content?

Yes

No

6. How do you use any of the social networking content for description of your assets or collections?

7. Does your system index user-supplied metadata?

- Yes
- No

8. Is user-supplied content searchable?

- Yes
- No

9. Do you perform any spell-checking of user content or de-duping of tags submitted by users (e.g. differences in capitalization or spacing, singular vs. plural, etc.)?

- Yes
- No

10. Do you make corrections to your existing metadata as a result of user contributions?

- Yes
- No

11. Do you have a formal process for making changes as a result of user-supplied content?

- Yes
- No

Page 9. Staff / Site Management

1. What roles do your staff serve on the site? [Check all that apply]

- Answer questions
- Participate in conversations
- Monitor site
- Nurture or further the community through outreach activities
- Add information & details to the site's content
- Staff perform no roles
- Other (please specify)

2. What staff play a role in managing your social metadata site? [Check all that apply]

- Information Technology Staff
- Interns
- Part of the duties of professional archivist/curator/librarian
- Primary responsibility of professional archivist/curator/librarian
- Volunteers trained by staff
- Other (please specify)

3. How much staff time (or other resources) are you committing to the creation and maintenance of your site?

- Less than 1 hour/week
- 2-5 hours/week
- 6-10 hours/week
- 11-19 hours/week
- More than 20 hours/week

4. Of the total time you spend on maintaining the site, what percentage of time goes into the following activities? Please round to the nearest 10%. Total should add up to 100%.

	Percentage of time spent
Maintenance of the site	<input type="text"/>
Adding new content	<input type="text"/>
Moderation	<input type="text"/>
Incorporating user generated content	<input type="text"/>
Other*	<input type="text"/>

*Other (please specify)

5. Is your site related to a single organization, or is it the result of a collaboration among two or more organizations?

- Single
- Collaboration among two or more organizations [Please indicate how administrative responsibilities for managing the site were negotiated]

6. How often do you or your colleagues add new content to your site?

- Daily
- Weekly
- Twice or more per month
- Monthly
- About every two months
- Three to five times a year
- Once or twice a year
- Haven't added new content since site launched

7. Do you think your project is sustainable?

- Yes
- No
- Unsure

Please comment:

8. Has your workflow changed since implementing your site?

- No
- Yes (Please outline the key points about how it has changed; you are not limited by the size of the box)

9. What types of policies or guidelines have you implemented? [Check all that apply]

- None - we have not implemented any policies.
- Acceptable community behavior and appropriate content
- Responsible re-use of the site's content
- Guidelines on format, controlled vocabulary, or general nature of contributions
- Users who violate policy may be blocked from site
- Trusted users may be whitelisted (contribute without moderation)
- Ownership of user-generated content is retained by site/institution
- Ownership of user-generated content is retained by and attributed to contributor
- Institution retains right to edit, repurpose, or remove user-generated content without notice
- Institution safeguards users' personal information
- Other (please specify; you are not limited by the size of the box)

Page 10. Staff/Site Management – Policies

1. Please provide the Web address(es) for your statements on policy, privacy, and terms of use below.

2. Are your site policies (check all that apply):

- extensions of existing institutional policies.
- new ones, created as the result of situations arising from the site.

Neither or other (please specify)

3. Have your policies had impact on staffing, workflow, or relationships with existing user communities and donors?

- Don't know
- No
- Yes (please briefly describe; you are not limited by the size of the box.)

Page 11. Design / Functionality

1. What algorithms are used to generate suggestions about similar content on the site to users?

- System does not generate suggestions
- What users are looking at
- Terms included in the user's searches
- Frequency of click patterns
- Other (please specify)

2. How are search results ranked? (Check all that apply)

- Search results are not ranked
- Relevance-ranked
- Most commented
- Most recent
- Other ranking order [please explain]

Page 12. Measures

1. Do you consider your site successful or unsuccessful?

- Successful
- Unsuccessful

Please tell us why (you are not limited by the size of the box)

2. What are your success criteria for your site? [Check all that apply]

- Adding new content
- Engaging new audiences
- Engaging audiences you did not expect
- Engaging and supporting existing audiences
- Gathering additional metadata about existing content
- None - have no success criteria
- Other (please specify)

3. Do you measure activity on your site?

- Yes
- No

Page 13. Measures continued [if answered yes to question 3 page 12]

1. What types of activities do you measure? [Check all that apply]

- Number of user accounts
- Number of unique visitors
- Number of groups formed
- Number of visits
- Most frequently accessed content
- Number of comments

- Number of tags
- Number of uploads
- Number of downloads
- Geography of users
- Domain IP
- User names
- Other (please specify)

2. How do you collect data? [Check all that apply]

- Web analytics (transaction logs, search logs)
- From hosted site
- Surveys
- Focus groups
- Registration information
- User-supplied content
- Other (please specify)

3. On average, how many visitors come to your site per month?

- Fewer than 250 per month
- 250 to 999 per month
- 1000 to 4999 per month
- 5000 to 9999 per month
- 10000 to 24999 per month
- 25000 to 49999 per month
- 50000 to 99999 per month
- Over 100000 per month

4. How many visitors contribute content to your site each month?

- Fewer than 100 per month
- 100 to 499 per month
- 500 to 1000 per month
- Over 1000 per month

Page 14. Thank you for completing the survey!

1. Thank you for completing the survey! If you have any questions or have suggestions of other people who you think should receive this survey, please contact Karen Smith-Yoshimura at smithyok@oclc.org. Your contact information will be kept confidential. We request the information below in case we need to clarify your responses.

Name:

Institution:

Email Address: