

# Blogger Perceptions on Digital Preservation

Carolyn Hank, Songphan Choemprayong, & Laura Sheble  
University of North Carolina at Chapel Hill  
Email: {hcarolyn, songphan, sheble} @email.unc.edu

**How do bloggers feel about digital preservation?** We asked bloggers that question in this study. Why? In the past decade, blogs emerged as established vehicles for communication and information dissemination. As of June 1, 2007, Technorati.com tracked over 84.5 million blogs. Presented here is preliminary analysis from an investigation into blogger perceptions on digital preservation issues, both for their personal blogs as well as the blogosphere in general.

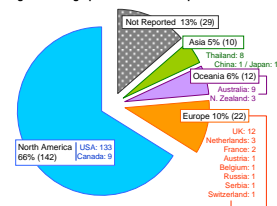
## Method

A web survey was distributed to a core group of 33 bloggers using a snowballing technique. The survey consisted of 40 to 46 questions across five subject areas: demographics, awareness, appraisal, impact, and investment. The survey was administered April 23-May 23, 2007. 215 surveys were completed.

## Demographics: Characteristics

**Respondents:** More males (55%) than females (45%) completed the survey. A large proportion were in their thirties (47%), but individuals represented all age ranges (less than 20 to over 60). Level of education was skewed towards highly educated individuals – 63.1% have completed a graduate or professional degree. A majority (66%) identified their location as North America, but Europe (10%), Oceania (6%) and Asia (5%) were also represented.

Figure 1. Geographic locations of respondents



**Authorship:** 47.6% of respondents post their single-author blog, while the same percent (47.6%) post to both a single-author and collectively authored blog (at least two authors). 4.8% post to collectively authored blogs only. For single-author blogs, 59% post to one blog; 23% to two; 11% to three; and 7% to four or more. For collectively authored blogs, 63% post to one blog; 25% to two; 11% to three; and 10% to four or more blogs.

**Experience:** The majority of respondents are quite experienced bloggers: 43% have blogged for 3 years or more, and 38% began blogging 1 to 3 years ago. Posting is, at minimum, a weekly activity for the majority (80%); 34% spend 1-2 hours per week blogging, and 29% spend 3-5 hours per week. 26% report posting at least 1x per day, and 30% report posting at least 3x per week. In addition, 90% of bloggers read other blogs daily, with 57% accessing other blogs at least 3x per day.

**Subjects:** Respondents blog about professional interests the most (73.5%), followed by technology (63.5%); life and personal experiences (57.8%); and educational/academic interests (54%).

**Audience:** The majority of bloggers identify colleagues and professional peers as their primary audience (63%), followed by friends (55.5%) and the general public (54%). About 42% indicate that they write for themselves as the "primary audience." 78.5% report that all of their posts are publicly available, 18.5% report some are password protected, while 3% require a password for the blog as a whole.

**Blog Elements:** Bloggers use the following elements: text (99.5%); links to other blogs (89%); photos (85.7%); comments (77.8%); tags/keywords (64.8%); RSS feeds (60.5%); blogroll (52.9%); permanent links (60%); video (40.5%); and audio (21.4%). Most include links to original source material (65.6%), verify facts to include in posts (62.7%), directly quote online resources (54.8%), and respond to posts and comments (50.2%). Occasionally, bloggers quote offline resources (45.6%). They rarely take content from other sources and remix it (35.9%), and half (49.5%) obtain permission to post copyrighted material.

## Acknowledgements

The study team thanks the Real Paul Jones, our faculty adviser, for his help in completing this study. This research was supported in part by a doctoral fellowship funded by the Institute of Museum and Library Services.

## Awareness: Persistence of Content

**Nearly 99% of bloggers report editing blog posts after initial posting** (often, sometimes, or rarely), and about 2 out of 5 (41.3%) report deleting blog posts after publication. Most often, posts are edited after publication to correct spelling or grammatical errors (48.73%). Bloggers are less likely to edit or delete posts that contain incorrect information or opinions that are no longer held.

Table 1. Reasons for post edit/delete

EDIT	Often		Never		DELETE
	Often	Never	Often	Never	
Spelling or grammatical errors	48.7%	1.0%	3.1%	56.9%	Correct erroneous information
Correct erroneous information	13.4%	13.9%	3.1%	66.2%	Update links
Update links	23.5%	12.8%	6.1%	68.2%	No longer interested in subject of post
Rework phrasing	23.5%	12.8%	6.1%	68.2%	Remove incorrect information
Remove incorrect information	4.2%	43.5%	4.6%	56.9%	Remove opinions no longer held
Remove opinions no longer held	0.0%	69.7%	4.6%	53.9%	Remove incorrect information

**About 23% have deleted an entire blog, 67% of these reporting intentional deletion.** Two prevailing reasons for intentional deletion are migration to another blog platform (43.9%) and loss of interest in subject matter (36.6%).

**In terms of the blog's skin, theme, layout and design, 80% of respondents have changed the look of their blogs on purpose, with 20% doing so 2 to 3 times per year and 11% changing appearance every few months.** 3.4% report that appearance was changed by their blog service providers.

Bloggers were asked to comment on three aspects of intentional blog preservation: duplicating posts, duplicating comments, and purposefully archiving all or parts of their blogs. 43.2% of respondents purposefully archive all or parts of their blogs. Additionally, nearly 70% of respondents have duplicated some of their blog posts and about 35% report duplicating some comments posted to their blogs. There is reporting discrepancies between these three actions, and the study team plans further investigation. A strong possibility for confusion was the construction of the survey questions.

Table 2. Extent of duplication

Posts	Percent	Comments
26.5%	All (100%)	11.9%
26.5%	Most (80-99%)	10.4%
10.3%	Majority (60-79%)	6.2%
6.6%	Half (40-59%)	1.0%
12.5%	Some (20-39%)	1.6%
17.6%	Few (1-19%)	5.2%

For those reporting purposeful archiving of all or parts of their blogs, bloggers report taking one or more of the following actions:

- Download and save to
  - Personal hard drive (67.4%)
  - Network hard drive (22.5%)
  - External media – e.g., CD-ROM, DVD (15.7%)
  - Use an archiving service (6.7%)
  - Print out blog (15.7%)
- Other – e.g., PANDORA, Rsync; MSWord file, etc. (18%)

**24.4% report archiving all or part of their blog according to a regular schedule.** Additionally, 7% do so in response to writing a post perceived as important, and 57% report archiving sporadically. 11.6% responded "other," and provided open text responses. For most of these respondents, archiving was done in response to an event (e.g., release of software upgrade).

## Appraisal: Attributes & Responsibility

The majority of bloggers identified themselves – blog authors/owners – as primarily responsible for preserving their own blogs (90.9%) as well as blogs in general (79.5%). Additionally, 20.5% identify blog service providers as primarily responsible in preserving blogs, for both their own blogs and the blogosphere-at-large. Perceptions on the role of libraries and/or archives in preserving blogs is substantially different. Only 4.5% thought libraries should preserve bloggers' own blogs. However, when asked in reference to blogs in general, the proportion of bloggers who believe that libraries and archives should take some responsibility grew to 30.7%.

Table 3. Responsibility for blog preservation

Their own blogs	Blog Authors/Owners	Blogs in general
90.9%	79.5%	79.5%
20.5%	20.5%	20.5%
10.2%	11.4%	11.4%
4.5%	30.7%	30.7%
4.5%	21.6%	21.6%
2.3%	9.1%	9.1%

The majority of bloggers agreed (36%) or strongly agreed (54.9%) that their own blogs should be preserved, while 11.1% strongly agreed that every blog should be preserved. For more information on what "types" of blogs were described as most "important to preserve," see <http://persistentblog.wordpress.com/>.

Table 4. Preservation perceptions - general

		Minutes		
		0-1	2-3	4-5
Should Preserve	Personal blog	70.9%	23.8%	5.3%
	Every blog	35.8%	27.0%	36.3%
	Every comment	31.4%	31.9%	36.7%
Should Not Preserve	All online content	28.2%	22.3%	49.5%
	Some blogs	44.7%	27.7%	27.7%
	Some comments	48.4%	31.3%	20.2%

The three characteristics selected as "very important" in determining blogs that merit preservation are 1) unique or innovative content of blog (52.2%); 2) cultural significance of blogger (46.5%); and 3) blog author's preservation preference (32%). The three characteristics bloggers selected as "very unimportant" are 1) size of blog (8.2%); 2) age of blog (6.1%); and 3) unique or innovative graphic attributes of blog (3.8%).

Table 5. Preservation appraisal - blog elements

Very important or important	Characteristic	Very unimportant or unimportant
88.6%	Unique/innovative content	3.3%
84.3%	Cultural significance (blogger)	3.8%
75.7%	Popularity	8.3%
70.1%	Number of links to blog	8.2%
67.2%	Number of regular readers	9.3%
66.7%	Number of hits	13.7%
65.2%	Preservation preference	10.5%
63.7%	Unique/innovative functionality	12.6%
54.7%	Unique/innovative graphic attributes	16.4%
52.8%	Popularity of (other) blogs linking to blog	17%
34.1%	Number of occasional readers	20.3%
32.2%	Age of blog	47.2%
23.1%	Size of blog	30.8%

Bloggers identified the following elements as most important to capture:

Text (95.4%)	Linked content (13.8%)
Images (53.9%)	Imported images – e.g., flickr (10.5%)
Audio (49.3%)	Imported tags – e.g., del.icio.us (10.5%)
Video (49.3%)	Other multimedia (43.4%)
Other multimedia (43.4%)	Comments (30.3%)
Comments (30.3%)	RSS feeds (10.5%)
Keywords – e.g., tags (28.3%)	Bookmarks – e.g., blogroll (9.9%)
Graphic layout/design (16.4%)	

## Impact: Risk Response

Bloggers responded to a delight scale to assess how they would feel if their blogs were deleted. With one exception, no bloggers indicated "positive feelings." Nearly 4 out of 5 bloggers would be either unhappy or miserable if their blogs were deleted.

Table 6. Reaction to loss/deletion

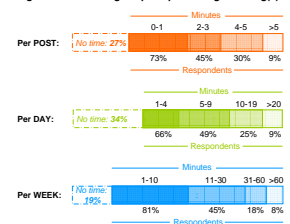
Miserable	Unhappy	Somewhat unhappy	Mixed	Somewhat pleased
34.8%	42%	14.5%	8.2%	0.5%

## Investment: Expending to Preserve

In response to how much time they would be willing to expend to preserve their blogs (per post, day, week, month and year), a portion of bloggers was unable to spend any time (0 minutes) to preserve their blogs – approximately, 26.6%, averaged across minutes per post, day, and week. Unsurprisingly, for those willing to expend time toward preservation, the majority prefer to spend a limited amount of time (e.g., per post, 43% willing to spend 3 minutes or less of their time).

When asked to react to a scenario where their blogs were deleted or "lost," many (39) indicated that, among the first things they'd do, is find out why and/or contact their service provider or blog technical support (62). Answers to this question illustrated the wide range of emotions bloggers expected to experience if a blog were lost, from "I shout, plead to my boss not to fire me!" to "I drink some coffee and do other hobby." For more responses to this question, see <http://persistentblog.wordpress.com/>.

Figure 2. Time willing to spend preserving own blog(s)



For expenditures of dollars, 53.3% are willing to expend monies (reported in US dollars) to preserve their own blogs. 2.4% of these bloggers report no limitation on amount willing to spend. The remainder, on an annual basis, are willing to spend:

- \$10.00 or less (26%)
- \$11 to \$25 (40.5%)
- \$26 to \$50 (15.5%)
- \$51 and greater (18%)

Additionally, 45.8% are willing to expend dollars for those blogs they find most important, and 29.2% are willing to expend dollars to preserve all other blogs.

## Wrap-Up

This work arose from a simple question: How do bloggers feel about digital preservation? Reflecting on our study design, one issue is our use of a snowballing technique for recruitment. Bloggers in this study are more professionally-oriented and include library and archival professionals, who are likely more interested in preservation issues than other groups. Future work would entail more effort to expand the core group further beyond ILS. One possible solution would be direct contact with service providers (e.g., blogger, wordpress) for recruitment. This study, and future efforts to expand participation, would be useful in defining the roles of different stakeholders (e.g. blog owners, blog providers, libraries and archives) in taking a step toward preserving this enormous content.

We consider this study a work-in-progress; "progress" because the findings show that, yes, bloggers are interested in the preservation of their content (but, not all blog content), and are generally willing participants in their blogs' persistent access and preservation, whether for public or private consumption. Future work will widen the net, so to speak, to better capture the sentiment of the blogosphere-at-large.

Copyright is held by the author/owner(s):  
Hank, C., Choemprayong, S., & Sheble, L.  
JCLD '07, June 18-23, 2007, Vancouver, British Columbia, Canada.

