The International Journal of Digital Curation
Volume 8, Issue 1 | 2013

Defining What Matters When Preserving Web-Based Personal Digital Collections: Listening to Bloggers

Ayoung Yoon,

PhD Student,

School of Information and Library Science,

University of North Carolina at Chapel Hill

Abstract

User-generated content (UGC) has become a part of personal digital collections on the Web, as such collections often contain personal memories, activities, thoughts and even profiles. With the increase in the creation of personal materials on the Web, the needs for archiving and preserving these materials are increasing, not only for the purpose of developing personal archives but also for the purpose of capturing social memory and tracking human traces in this era. Using both survey and interview methods, this study investigated blogs, one popular type of UGC, and analyzed travel bloggers' perceptions of the value of blogs and the elements of blogs that are important for preservation. The study respondents found personal and sentimental value (e.g., a way to express themselves, a way to keep personal memories and thoughts, and a way to maintain a record for their family) to be the most important reason for preserving blogs, followed by informational value and cultural/historical value. Sharing also appeared as one of the values that respondents found in their blogs. The respondents reported that self-created blog posts (content) and information related to the blog posts (context) are more important to preserve than some other elements (behavior and appearance). Integrating what bloggers consider as most valuable and what archivists think are worth preserving may be an important step when collecting personal blogs.

Introduction

As the Web has become a normal channel of content publishing for individuals, more and more personal materials have been created and disseminated through the Web. Anyone who has access to the Internet can create user-generated content (UGC) using Web-based tools such as blogs, YouTube, Facebook and Twitter. The 2007 Organization for Economic Co-operation and Development (OECD) report on the participatory nature of the Web defines UGC as online content that is produced by amateur creators and made publicly available on the Internet. This content reflects a "certain amount of creative effort," and is "created outside of professional routines and practices" (Wunsch-Vincent & Vickery, 2007). UGC is a collection of cultural narratives from ordinary people; the narratives can be a part of personal digital collections, as they often contain material such as their personal thoughts and experiences. Because individual materials and stories are increasingly reflected on the Web and diffused through the Internet, a "new archival calling" – as Cox (2009) called it, in acknowledgement of the importance of personal archives in society and the partnership between the public and archivists – is emerging as a way to preserve human traces of this era.

Recognizing the importance of preserving individual materials and stories on the Web, this study targets blogs – particularly travel themed blogs, a popular type of UGC – and aims to understand those bloggers' perceptions of preservation and their actual preservation behaviors. To do so, the following questions were addressed. First, the study asked what values bloggers give to their blogs, as value is foundational in archival appraisal theory for preservation. Second, the study asked how bloggers define what is important to preserve. Using the concept of significant properties, the second question addressed the actual elements of blogs that bloggers want to preserve, which reflects what they perceive as the essential elements of blogs. Lastly, this study investigated what actual blog preservation behaviors bloggers use to preserve these essential elements. Learning actual preservation strategies revealed whether bloggers' value assessment of blogs and their choice of elements can be reflected in real preservation behaviors, and how.

Understanding bloggers' perspectives and integrating them into preservation efforts could facilitate management and preservation of blogs for personal purposes, as it can be helpful to bloggers who are developing personal digital archives. It would also help archivists to develop new approaches for appraisal of personal collections, as personal materials require different concepts and treatment by archivists because these materials are not typically generated as transactional or public activities (Hobbs, 2001). Additionally, it is beneficial for programmatic blog preservation activities to develop preservation tools for repositories and adopt these tools into workflows.

Preserving Personal Digital Collections on the Web

The Value of Personal Digital Collections and Blogs

Beagrie (2005) defines a "personal digital collection" as a part of an informal "personal archive" of records to "distinguish informal, diverse, and expanding collections accumulated and maintained by individuals." Similarly, Williams, John,

and Rowland (2009) use "personal digital archives" to refer to "informal, diverse, and expanding memory collections created or acquired and accumulated and maintained by individuals in the course of their personal lives, and belonging to them, rather than to their institutions or other places of work." Thus, personal digital collections can be understood as content created and maintained by individuals that reflects their personal lives.

Researchers have been arguing about the need for archiving and preserving personal digital collections for future research (Cunningham, 1994; McKemmish, 1996; Paquet, 2000). Cunningham (1994) pointed out that personal records have been overlooked in the past, and discussed the roles of archivists who manage personal electronic records. McKemmish (1996) explored the nature of personal record keeping and its role in constituting a part of collective memory and cultural identity, arguing how "evidence of me" becomes "evidence of us."

A series of projects investigated this issue further. One example is the W.D. Hamilton Archives from the British Library in 2000, which examined Professor Hamilton's electronic manuscripts, electronic messages and other materials (Summers & John, 2001). The Personal ARchives Accessible in DIGital Media (PARADIGM[1]) project by the University of Oxford and University of Manchester explored archivists' roles in selecting, acquiring, processing, storing and preserving individuals' digital collections using the papers of contemporary British politicians as a test bed (Thomas & Martin, 2006). More recently, the Digital Lives[2] project at the British Library addressed curatorial issues of personal digital collections (Williams, Dean, Rowlands & John, 2008).

UGC has become a significant part of personal digital collections on the Web, which often contain personal memories, activities, thoughts and profiles. Blogs are particularly interesting because the majority of them are personal journals or diaries (Herring, Scheidt, Wright & Bonus, 2005) that include potentially revealing content, although they exist in the public realm unless action is taken to secure access. Because personal digital collections have the interactive characteristics of social networking sites, personal records in UGC are generally shared more than similar types of offline content (Lee, 2011). At the same time, unlike traditional diaries, blogs also enable users to incorporate different media, such as photographs, audio and video. Thus, blogs are also contributing to or reshaping documentary heritage by expanding the concept of the diary (Cox, 2009).

Blogs that have the characteristics of personal journals may be seen to have particular value in a cultural context. Cultural historians have stated that personal profiles are evidence of "social characteristics and changes" (Nesmith, 1982) as well as "a window into the past and an exceptional source of information for the life and times of a particular diarist" (O'Sullivan, 2005). Because individual memory can only be recalled in the social framework within which it is constructed (Halbwachs, 1992), collecting and preserving personal materials on the Web can be a valuable way of preserving our culture and social memory. Lee (2011) also argued that these personal collections may become a matter of "capturing the 'externalized me' that individuals

---

[1] PARADIGM: http://www.paradigm.ac.uk/

[2] Digital Lives Project:
http://web.archive.org/web/20110208202601/http://www.bl.uk/digital-lives/index.html.

have left in various online locations," and in order to track human traces on the Web and to understand individual lives for future research, it is necessary to give more attention to preserving personal digital collections.

Although archives that preserve personal papers and collections have long been a cultural institution, Hobbs (2001) said that most archival theory does not address personal records, mainly because archival theory was built on the model of organizational and governmental practices. Thus, Hobbs argued the need for developing new archival approaches for personal collections and papers that emphasize an "individual and idiosyncratic" rather than the current "corporate and collective" approach. As personal collections are "exporting their own life" (Hobbs, 2001), listening to creators – bloggers, in this study – can be one way of broadening archival approaches to collecting and preserving personal collections.

Self-Defined Value of Personal Archives

Value has different meaning in different contexts. The Society of American Archivists glossary defines *value* as "the usefulness, significance, or worth of something to an individual or organization"[3] in archives and records context. Archivists have addressed personal collections as a part of social memories and history, and discussed the value of personal papers. How individuals value their personal archives (both physical and digital) has been actively discussed in the field of personal information management and human-computer interactions (e.g., Kaye et al., 2006; Marshall, Bly & Brun-Cottan, 2006; Whittaker & Hirschberg, 2001; Williams et al., 2008). In particular, Marshall et al. (2006) identified possible factors that influence individuals' value judgment, including personal labor, creativity, emotional impact of a personal item, and replaceability (e.g., how hard it is to replace), all of which aid in determining how individuals assess their belongings.

Other studies have suggested different values and reasons for keeping personal archives. Those values not only refer to "value of something" (e.g., artifactual value and legacy value) but also "value of doing" (e.g., value of actions, such as referencing and sharing). Thus, some values are more related to the archival functions, while other values relate more to sentimental or emotional reasons. For instance, referencing information (Whittaker & Hirschberg, 2001; Kaye et al., 2006) and sharing it (Kaye et al., 2006) are values that individuals have been found to reflect in relation to the functions of personal archives. Regarding sentimental or emotional values, Kaye et al. (2006) found that personal archives hold a legacy value, which means they maintain individual personality or important aspects of life's work. Williams et al. (2008) found that individuals look at their archives as a reflection of their life's work, and Kirk & Sellen (2008) found connecting with the past and framing the family as one of six values of home archiving. Other values discussed include constructing identity or defining self (Kaye et al., 2006; Kirk & Sellen, 2008), and preservation from fear of loss, for example (Whittaker & Hirschberg, 2001; Kaye et al., 2006). The context of each study is slightly different (e.g., personal papers in office work, or personal archiving in physical and virtual spaces or at home), but these findings can all provide useful insights about bloggers' perceived values of their blogs.

---

[3] Society of American Archivists (SAA) glossary definition for "value": http://www2.archivists.org/glossary/terms/v/value

Blog Preservation

The issues of blog preservation were first addressed with the Web archiving approach (e.g., the Wayback Machine from Internet Archives), but discussions about blog preservation needs, actions and applications have recently been expanding. Starting in the mid-2000s, several articles called for particular attention to blog preservation (Entlich, 2004; O'Sullivan, 2005; Paulus, 2007), and in 2007, Hank, Sheble and Choemprayong (2007) first surveyed bloggers to learn their perceptions about preservation. Recently, more discussions have emerged within institutions and organizations in order to assess how institutions can preserve institutional, professional or scholarly blogs (Davis, 2009; Reich, 2009; Kelly, 2010; Kelly & Guy, 2010). Several projects were also initiated based on the need to develop blog preservation tools. For example, the ArchivePress project argued for the development of an innovative blog archiving application to help institutions with blog management, and developed a demonstrator tool (Pennock & Davis, 2009). BlogForever was initiated to provide a solution for blog archiving and digital preservation by ensuring authenticity, integrity, completeness, usability and long-term blog accessibility (Banos, Arampatzis, Cristea & Tero, 2010).

While attention to blog preservation has been growing, a gap remains between empirical evidence on content creators' preservation preferences and perceptions and the literature on the subject. By closely investigating content creators' perspectives, this study helps to fill that gap.

Significant Properties of Blogs

Understanding that the potential value of personal digital collections and efforts to preserve them are significant, the question arises of what aspects of this content should be preserved. Defining the "essential characteristics" or "significant properties" of digital objects is one mechanism for identifying the characteristics that must be preserved for these digital objects to remain meaningful and usable. Several projects have expanded the discussion regarding the essential characteristics of digital objects for long-term preservation. The CEDARS[4] project, which first used the term "significant properties," argued that it is important to retain "the level of content and functionality" needed to ensure "the object is accessible to the agreed level" (Cedars Project Team, 2001), and the CAMiLEON[5] project identified "properties of digital objects that affect their quality, usability, rendering, and behavior" (Hedstrom & Lee, 2002). Later, the InSPECT[6] project defined significant properties as: "the characteristics of an information object that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the object" (Wilson, 2007).

In order to identify elements of significant properties of digital objects, the InSPECT project categorized five components:

---

[4] Curl Exemplars in the Digital ArchiveS (CEDARS).

[5] Creative Archiving in the Michigan & Leeds: Emulating the Old on the New (CAMiLEON)

[6] InSPECT project: http://www.significantproperties.org.uk/

1. Content (e.g., text and images);

2. Context (e.g., who, when, and why);

3. Appearance (e.g., font, color, and layout);

4. Structure (e.g., pagination, headings, and embedded files);

5. Behavior (e.g., hypertext links, updating calculations, and active links).

The concept of significant properties has been applied to different types of digital objects, such as emails, vector images, audio recordings, raster images and moving images (Knight, 2010a, 2010b; Coyne, Duce, Hopgood, Mallen & Stapleton, 2007; Montague, 2010; Coyne & Stapleton, 2008), but not to blogs. Although defining significant properties of blogs might present challenges due to their dynamic nature, this effort would help to clarify the essential elements for preservation from bloggers' perspectives.

Method

This study involved two data collection methods: an online survey and a follow-up semi-structured interview. Survey questions were developed based on survey instruments designed in previous research (Hank et al., 2007). The survey included a total of 19 questions asking about basic blog activities, bloggers' value assessments, elements of significant properties identified by the InSPECT project, and self-reported preservation behaviors. Blog performance issues were addressed in the current study. Thirty-minute phone interviews were conducted with a subset of the survey respondents in order to gain in-depth perspectives of blog preservation and preservation behaviors.

As a way to limit potential respondents to personal bloggers and omit those who pursue business objectives, this study focused on bloggers who have written about the theme of travel, although their blogs were not exclusively limited to that subject. This theme was chosen, in part, because travel is one of the most popular themes among bloggers (Technorati, 2011). In addition, travel blogs often contain various forms of information, such as text, photographs and videos, and bloggers write about travel information as well as personal experiences related to travel.

Potential survey respondents were identified by looking at popular tag clouds generated by WordPress, a well-known blog hosting service in the United States. Respondents were invited to participate in the survey through two different forms of invitations: e-mail (if potential respondents shared their e-mail addresses on their blogs) and direct reply to potential respondents' blogs (if they did not share contact information on their blogs). The online survey was first distributed to bloggers in mid-March 2011 and opened for a month. Three hundred and seventy email invitations and 200 direct blog replies were sent to bloggers, and 135 bloggers responded to the survey (a response rate of 23.68%). After the survey was completed, the author examined respondents' blogs and screened unqualified blogs (e.g., commercial travel agencies' blogs or non-travel-related blogs) and excluded them from the results. Incomplete responses were also excluded. As a result, among the 135 bloggers who participated in the online survey, 114 responses were analyzed.

Interview participants were recruited by email invitation using contact information provided by survey respondents, as the last survey question asked whether the respondent was willing to be contacted for future study or interviews. Nineteen bloggers volunteered for interviews, but eight bloggers were actually interviewed for various reasons (e.g., withdrawal or scheduling issues). All interviews were conducted via telephone.

Raw survey data was re-assessed for its completeness and exported to SPSS, software for statistical analysis, for further analysis. Open-ended responses were analyzed using Excel. Interview data were fully transcribed and analyzed using the TAMS Analyzer[7] based on themes that were inferred from the survey responses. Examples of themes from interview data included the purpose of blogging, the value of blogs, preservation definitions, preservation elements, preservation behaviors, and the reasons for preservation behaviors.

Findings

Demographics

The majority of survey respondents were female (71.1%) and the majority of respondents were aged 20-30 (see Table 1). Most survey respondents (82.4%) held degrees from four-year colleges (43%) or higher-level programs (master's degree: 27.2%; doctoral degree: 9.6%; and other professional degree, e.g., JD/MD: 2.6%). The majority of respondents were from English-speaking countries, and more than half were from the U.S. (57%). Other countries of origin included the U.K. (12.3%), Canada (7.9%), and Australia (3.5).

Age	Percentage	Frequency
Younger than 19	1.8	2
20-29	47.4	54
30-39	22.8	26
40-49	13.2	15
50-59	9.6	11
Older than 60	2.6	3

Table 1. Age of survey respondents (n=114).

Interview participants also included more females (5) than males (3). The ages of participants varied: two were in their 20s, three in their 30s, two in their 40s, and one in her 50s. Occupations included student (both undergraduate and graduate), volunteer, homemaker, professor and businessperson.

---

[7] TAMS Analyzer: http://tamsys.sourceforge.net/

Purpose of Blog Preservation: Why Blogs Are Worth Preserving, and for Whom?

Value of blogs

A survey question asked respondents to what extent they agreed or disagreed with statements regarding why they value their blog posts, allowing for free-text responses in case they held values not represented by multiple-choice answers. The survey respondents all reported that they held at least some value for their blogs and disagreed with the statement "I don't see any value" (Table 2).

	Mean	Standard Deviation
My blog posts are my personal diaries that hold my history and memories.	4.28	0.88
My blog posts provide useful or valuable information to others.	4.19	0.75
My blog posts are culturally or historically significant.	3.44	1.07
I don't see any value in my blog posts.	1.48	0.85

Table 2. Extent to which respondents agree or disagree with each statement regarding the value of their blog posts. 5 = Strongly agree; 1 = Strongly disagree (n=114).

Among different kinds of values, respondents found personal value in their blogs, which represent their personal memories and their personal or family history. From the analysis of text responses in the survey and interview data, three subsets of personal values emerged: a way to express themselves, a way to keep personal memories and thoughts, and a way to keep a record for their family. For some, writing blog posts means reflecting on their life on the Web. Like diaries, these stories or records are a way to document and memorialize special events in the bloggers' lives or trivial activities in everyday life. In the case of IU07, blogs are valuable (future) family records for children:

> 'Also for me, the value … is for my family and specifically my children to be able to see thoughts that I am preserving or I am thinking about. My children are very young, I'm a young parent, but for that reason I want them to know what I was doing in my life if something were to happen to me or in 50 years, they take up the blog book, they can see somewhat of the history of what their father was involved in, in life, what he considered valuable, what he considered worthy of his calling and time.' (Respondent IU07).

Respondents also perceived their blogs as an information source for others. Both survey respondents and interview participants felt that blog posts were useful, whether they were personal stories (e.g., evaluations or emotions) or objective information (e.g., location name, price or other visitor information), as either type of blog post could strike readers. In particular, two interview participants found the value of a blog in its ability to serve as a unique source of information (e.g., sharing life as an amateur

surfer or travel volunteer). For IU08, the value of a blog was not only in sharing information, but also in sharing information that was hard to find anywhere else:

> 'My blog, specifically, is about volunteering while you're traveling. What I've discovered in writing it is that there really isn't another site exactly like mine and so it seems like there is a need for this type of information out there.' (Respondent IU08).

Although survey respondents did not agree as strongly that their blogs held cultural or historical value compared to personal or information value, some acknowledged in their text responses the importance of personal history to society for its cultural and historical value. Several interview participants also mentioned the significance of keeping unique individual histories for the purpose of maintaining cultural diversity or preserving different ideas as a representative record of a particular time. For example:

> "[J]ust personal history or personal experiences that are unique to our culture, whether it would be like now, maybe there is some soldier in Afghanistan keeping a blog and I think those experiences are valuable.' (Respondent IU01).

> '[B]logs in general is not just a sort of satisfying the ego of the writer himself, but I think it's socially relevant and it captures sort of the views and ideas of the populous and related to a particular time.' (Respondent IU06).

Sharing also appeared to be valuable to survey respondents. In their text responses, more than half (30 out of 48) said they appreciated how blogs allow them to experience interconnectedness with others, such as communicating long-distance with family, friends and other bloggers. Of this group, about half identified "sharing experience or ideas with others" as their main purpose for blogging. For them, whether a certain blog post has personal, informational or historical value, sharing usually exists as an underlying value.

Later in the survey, the majority of survey respondents (74.8%) also expressed interest in making blog posts available for future use. Respondents repeatedly expressed values they found in their content as reasons for that decision – namely, historical records for future generations and useful information for others. Interestingly, a number of respondents wanted to donate their blogs to archives or other institutions for future use as a way of guaranteeing permanent access or preservation. On the other hand, among those who did not want to make their blogs available for the future, most were concerned about copyright or controlling their own content. Some wanted to know more about what was meant by "future use" and wanted to make sure that "future use" would not distort the original meaning of their content.

Intended audience

As sharing is one of the major functions of blogs as well as one of the major values, respondents answered that they expected their blog posts to be read by others. Friends were the most commonly expected group to read blogs (81.6%), followed by family (68.8%), the general public (67%), other bloggers (54%), colleagues (36%), and professional network (16%), including themselves (34%).

The intended audience group influenced the length of time that participants wished for the audience group to have access to their blogs. About half wanted to provide access to their blogs after their lifetime (see Table 3). Except for the blogger themselves, they wanted to provide permanent access to their family and friends, who probably have closer relationships with them than other audiences and are regarded as the most significant audience groups. Public access ranked second highest. This result coincides with the aforementioned finding regarding which audiences the respondents expected to reach. Respondents considered family, friends and the general public more targeted audiences than other groups.

	Permanently		During my lifetime		Some time but less than my lifetime		Not for any time	
	%	Freq	%	Freq	%	Freq	%	Freq
For my personal access	49.1	56	43	49	7.9	9	0	0
For family and friend's access	53.5	61	28.9	33	16.7	19	0.9	1
For my colleagues' access	31.6	36	21.1	24	30.7	35	13.2	15
For professional network's access	33.3	38	19.3	22	28.9	33	14.9	17
For public access	41.2	47	27.2	31	28.9	33	2.6	3

Table 3. How long respondents wanted their blog posts to be preserved (n=114).

Blog Elements for Preservation: What Elements Are Important to Preserve?

Respondents who answered that they want to preserve their blogs (n=113) were asked different questions regarding how they would decide what to preserve. All respondents said they want to preserve at least a certain amount of their blog posts, and none of them answered they do not want to preserve any of their blog posts. As can be seen from Table 4, more than 60% of respondents wanted to preserve all of their blog posts.

	Percentage	Frequency
All of my posts (100%)	60.5	69
Most of them (80-99%)	18.4	21
A majority (60-79%)	13.2	15
About half (40-59%)	5.3	6
Some (20-39%)	1.8	2

Table 4. Portions of their blog posts that respondents would like to preserve (n=113).

Common reasons for partial preservation, identified from the text responses of the survey, were that some of the blog posts were "not relevant in the long term," "not as important or useful as other posts," or "not original."

> 'I'd like the chance to go through and delete any posts that are egregiously stupid.' (Respondent Q12SU19).

> 'Many posts are links to other sites, news items, etc. I would preserve those only if they were truly significant. Otherwise, only original materials.' (Respondent Q12SU16).

Respondents reported varying opinions on which blog elements were worth preserving, such as user generated content (any type of content respondents created), any other content generated by others, design related elements (e.g., font, color, etc.), blog structuring elements (e.g., tags), and blog analytic elements. Respondents gave priority to content created by themselves for preservation (see Table 5). The top four elements respondents perceived as important were all forms of self-created content, regardless of the format of the content, and copied content had least importance for preservation. The different mean values among different formats of content may possibly be explained by respondents' preferences, as not all respondents utilized all blog elements. For instance, one respondent said, "My blog, it's the words and pictures that I take, like those pictures are personal pictures. So I think those are the big two things that are valuable to me" (Respondent IU04).

	Mean	Standard Deviation
Photos that I owned and posted	4.54	0.88
Text that I wrote	4.53	0.74
Video that I owned and posted	3.92	1.13
Audio that I owned and posted	3.74	1.4
Order or chronology	3.73	1.1
Date and time of posts created	3.72	1
Comments from others	3.72	0.9
Hyperlinks and other references to content outside my blog	3.6	1
Any visitor information for my blog (e.g., UV, name of visitors, etc.)	3.32	1.03
Images other than photos that I owned or posted	3.17	1.09
Layout of content (e.g., font, color or other appearance)	3.11	1.17
Information about friends who have connected to my blog	3.11	1
RSS feeds	2.84	0.96
Tags and categories	2.84	1
Any type of content that I copied and not owned	2.84	1.03

Table 5. Respondent ratings for the importance of individual blog elements to the preservation of the blog. 5 = Very important; 1 = Very unimportant (n=114).

Respondents perceived a certain structural element (chronology) as more important than other structural elements, including tags and categories, although all help to organize blogs. In blogs, posts typically appear in chronological order, with the newest post first. This order can be important because it serves as a reference point when bloggers want to go back to older posts. Respondent IU06 said:

> 'I think it's mostly important to preserve the chronology of the blog. Because … I mean when you're blogging, you're generally writing about events that occurred at that time you're writing. So the chronology is important because it establishes sort of a thread of continuity between your posts and also with things that are going on outside. And creating some sort of pseudo relevance to larger events.' (Respondent IU06).

Comments from others appeared to be contradictory between the survey results and interview data. Comments appeared to be an important element in the survey. However, although most interview participants (Respondents IU02–IU07) agreed that it is nice to receive comments and emotional support from others about what they have been doing, and they expressed their appreciation for comments, they did not feel very concerned about preserving comments.

> 'I do value them [comments] because it's like sort of people outside of my realm of friends that are actually reading my blog and actually think that I have something interesting to say or interested with what I'm writing. … For preservation, I don't think they're as important to me. I mean, it's nice to have and if I can preserve those, that's great, because it shows that there was interest in this specific blog entry, but if their comments aren't preserved, it's not like a huge loss. I don't think.' (Respondent IU04).

Respondents IU03 and IU07 explained that their reason for not strongly wishing to preserve comments is the limited number of comments they have received. IU03 said, "I am not as very concerned about the comments, because I think we don't have like hundreds and hundreds of comments," and similarly, IU07 said, "Primarily because I get so few comments. My blog is not one that traffics a lot of comments and even looking at it today and I've had about 40,000 hits, but I've only had 300 comments."

Respondents' attitude towards visual appearance, including color, font and other presentation features of blogs, also appeared to differ between survey and interview data. While visual appearance was perceived as having neutral importance (M=3.11) in the survey, all interview participants expressed their desire to preserve the visual appearance of their blogs. They wanted to make their blog "visually appealing" (Respondent IU01), and "work hard to get them laid out so they look good. [Laugh]" (Respondent IU05). In addition, IU05 said: "and not everybody has that ability [to make posts visually appealing] so I'd like to keep that."

Bloggers were also asked which versions of posts they wanted to preserve if they had edited the blog posts for any reason. Most respondents (86.8%) wanted the final version to be preserved, and only a small percentage wanted to preserve all versions (6.1%) or any version that was important (5.3%). This trend can be explained by bloggers' editing behavior. Respondents edited blog posts most frequently for minor

changes, such as spelling or grammatical errors (80%), or to correct technical problems such as formatting or publishing errors (72%). Changes in content occurred rarely, for situations like removing information that was no longer seen as valid (21%).

Preservation Behavior

Bloggers put great effort into creating blogs posts, which reflects the characterization of UGC as an outcome of a "certain amount of creative effort" (Wunsch-Vincent & Vickery, 2007). Interview participants expressed their concerns about unexpected damage to blogs, and used negative emotional terms such as being "sad and upset" (Respondents IU03 and IU07), "shocked" (Respondent IU04), "disastrous" (Respondent IU05), "freaked-out" (Respondent IU06), or "crushed" (Respondent IU08) to describe how they would feel if their blog was damaged. Participants, thus, had some preservation strategies to maintain their blogs or the contents of their blogs, although these strategies were often limited to minimal; for instance, respondents were only able to keep content but not other elements of blogs.

Survey respondents were asked about their levels of experience with preservation of their personal blog, defined as "any act of maintaining information in understandable form over time for ensuring its accessibility." More than half of the respondents (63.2%) answered that they have considered preserving their blogs. Among them, 71.1% have actually tried to preserve their blogs in some way, while 28.9% have not tried.

Different methods were chosen for preservation, but saving copied blog posts in the media they chose was the most common means of preservation among respondents (35.9%). When saving blog posts, they used personal hard drives (27.2%), network hard drives (2.6%), or external media such as CD-ROMs (6.1%). Other methods include printing blogs as paper books (4.4%), using commercial archival services (2.7%), and donating them to organizations (1.8%).

Interview participants were asked to elaborate their blog preservation strategy. The participants described three methods of preservation:

1. Only saving text and pictures, saving text as a Microsoft Word file and saving pictures separately (Respondents IU02, IU06);

2. Using commercial software (Respondent IU05);

3. Printing blog posts on a regular basis (Respondent IU07).

For the first case, IU02 and IU05 only saved what they considered the most important content of their blog posts (usually text and pictures) onto hard drives. Interestingly, both participants wrote text in Word documents first rather than writing directly in the blog's template, then copied the text into the template and added their chosen pictures. Thus, they already had Word files containing the contents of their original blog postings. Pictures were also already saved onto a hard drive or CDs. Since they already had pictures, and created Word files later, text and pictures were managed separately rather than in one document as blog posts. IU02 felt that having all of these materials together in the same way that the original blog posts appeared would make more sense, but IU02 also thought this would take more time, saying,

"I'm not doing it on one file. I mean I guess if I had more time, I would maybe do it all together because then it would be done. But it would take a lot more time." IU06 was willing to save blog posts as one file that had both text and pictures, but she did not know how to do it.

> 'Yeah, they're [text and pictures] all separate. … I can't change the format of the post and save it as it is onto my local machine.' (Respondent IU06).

For the second case, IU05 used a commercial service to preserve not only content but also other elements, including layout and design. Although IU05 had his own back-up system, copying text from "important" posts into Word documents and saving them, he wanted to preserve the original look and feel of the blog. This commercial service copies the entire blog and moves it to another Web server so users can access it through their own servers in the future.

> 'Well, I just pay for them to back it up. … You can just pay them to back it up, you can't physically do it yourself. I'm actually pretty technical, and I like technology but it bores me a little bit, and I have to worry about the backups. So I just pay them $10 a year, and they supposedly back it up. So I'm hoping all the files are there. I copy and paste most of them, text wise. But I really like the preservation of the way it's laid out, I really like the way it looks so it would be nice to be able to have it preserved exactly like it's laid out originally. The layout and look and feel matters a lot to me.' (Respondent IU05).

Lastly, IU07 printed all posts every two years using a commercial blog print service and turned them into a book. A printing service enables the blogger to have a copy that "looks exactly the same as the posts" although this method has some limitations such as not being able to print video or preserve live links. Still, IU07 believed this method is a good preservation strategy:

> 'It's basically a service that you basically go online, tell them what your domain name is, you give them the log in and the password, and then you give them a time parameter to this date to this date. And then basically, it will just copy them. It's everything that it copies; if it's a photo it copies photo, if it's text it copies text, a link copies a link. They're not live, but it will copy it out. The only thing it doesn't copy is video and then it doesn't really give you a way to assort that. All you can do is delete the blog entries that you do not want. … But it really is just for me and I have two sons and it becomes, in a way, a way for them in years down the road to kind of see what their father was thinking about, what I was working on kind of a scrapbook if you will.' (Respondent IU07).

As discussed, while survey results indicate that survey respondents generally consider design elements and layout less important than the content they create, these two interview participants wanted to preserve their blogs with the design elements intact.

Discussion

This study investigated the value of blogs as perceived by travel bloggers, important blog elements for preservation, and preservation behaviors of bloggers. Findings about perceived value suggested that bloggers perceived the personal value (e.g., representing a way to express themselves, a way to keep personal memories and thoughts, and a way to keep a record for their family) of their blogs as most important, followed by informational value and cultural/historical value. As blogs can be seen as one form of personal collections, it is not surprising that respondents strongly agreed that their blogs have personal and sentimental value. Similar to the findings in previous studies (e.g., Kaye et al., 2006), this study illustrated that blogs portray a personal legacy, hold memories and serve as a means of family framing.

While personal value appeared strongly among both survey and interview participants, it is worth noting that participants also understood their blogs' broader value to society, such as their cultural or historical value, which archivists often find in personal materials in addition to the individuals' personal and sentimental value. Often, these values were the reasons why respondents were interested in preservation and willing to make their blog posts available for others in the future.

Sharing is one of the major functions of blogs, as it allows bloggers to interact with readers; sharing also appeared as a value itself, as Kaye et al. (2006) found. The value of sharing might or might not apply to other types of personal digital collections, but as blogs have the innate characteristic of interconnectedness, sharing can be ascribed to them as a value.

Values and audience may influence the preservation strategies that individual bloggers would take. Those who consider their blogs as personal diaries or family records, might to want to preserve blogs for themselves and for family, while those who see more informational or cultural value in them seem to want to preserve their blogs for others. For example, respondent IU07 who wanted to preserve his blog for his young children in order to provide details about their parent's life – and thus was interested in preserving it for personal and family use – decided to create a printed book of blog posts. Printing has its own limitations and was not a popular method among survey respondents, but IU07 thought it was a good strategy for the purpose of leaving his blogs to his children.

Regarding the components of blogs that are seen as important for preservation, there is a tendency to prioritize self-created content as opposed to copied (or imported) content from others. Among self-created content, text and pictures are particularly significant because they are the most dominant content formats. Other blog elements, such as copied content regardless of format, visitor information, tags, hyperlinks, and comments, are less significant than content created by bloggers themselves. Although respondents often expressed their appreciation for this information (e.g., visitor information and comments) and technical functionality (e.g., tags and hyperlinks) provided by hosting services, it seems that what they assign higher priority are their stories and related information, not others' stories. One apparent example is blog comments. While respondents of this study enjoyed the attention from others that appeared through comments on their blogs, several

interview participants thought "it's not a huge loss" if those comments are not preserved.

One interesting finding is the different attitudes of participants between the survey and interview data regarding the visual appearance or design elements of blogs. In the survey, design elements of blogs appeared of less concern, but interview participants expressed their desire to keep the original look and feel of their blog intact because they valued visual representation. Reasons for this inconsistency between the survey results and interview data do not explicitly appear in this study. Perhaps interview participants may have found more value in design and layout than others did. Another possible inference is the gap between perceived value and capability of preserving valuable elements. Content would be still of central importance and would have higher priority for preservation, but bloggers might like to preserve all design elements if they knew how to do it. Even among interview participants who showed interest in preserving design elements, only two of them who had the capability actually preserved their blogs with design elements, using different strategies. Others who only saved text and pictures of their blog posts simply did not use other approaches that could possibly help them to preserve design elements.

Returning to the five concepts of significant properties from the InSPECT project – content, context, appearance, structure, and behavior – it is possible to conclude the following based on the survey results:

- Content and context are the most important blog aspects for preservation for this study population, as respondents definitely expressed their willingness to preserve their own self-created content, mostly represented by text and pictures, as well as information about this self-created content, such as the date of creation and author information.

- Some structural elements, such as the chronology of blogs, are important because they add context, but other structural elements, such as tags and categories, have less importance to the respondents.

- Appearance and behavior do not seem to be significant considerations for the respondents, as a majority of them did not express much interest in preserving font, color, layout, hypertext links, or records of updated information on blog posts.

While the survey respondents provided much information on their perspectives about blog preservation, the skewed demographics of the respondents is a limitation of the study. The demographics of survey respondents generally aligned with the demographics of bloggers—young and educated—reported by Technorati (2011), but survey respondents of this study were female-dominant while the Technorati report found that three-fifths of bloggers are male. This might stem from the higher response rate of women in survey research (e.g., Sax, Gilmartin & Bryant, 2003), or it may result from the limitation of blog topics to travel and individual lives. In addition, the respondents of this study were identified from one major blog hosting service, WordPress, which made it possible to access the target population, but this approach might not allow bloggers who use other services to participate in the study. More studies – possibly employing different recruitment techniques – that address different types of blogs using different hosting services would complement the results of this study.

Conclusions

The nature of a blogs' content is personal and the creator's perspectives need to be incorporated into decisions regarding appraisal and preservation. The values that bloggers assign to their content may or may not align with the values important to collections managers or archivists as they collect and preserve personal blogs. For example, the audience defined by bloggers might be different from the intended users or designated communities defined by archivists or collection managers. Depending on the archive's intended uses, some elements that bloggers tend to disregard, such as comments, might be important to keep. Regardless, integrating what bloggers consider as most valuable and what archivists think are worth preserving would be an important step when collecting personal blogs; this would develop new archival approaches for collecting and preserving personal materials that are "individual and idiosyncratic" (Hobbs, 2011).

Acknowledgements

References

Banos, V., Arampatzis, S., Cristea, A. & Trochidis, I. (2010). BlogForever project presentation. Retrieved from http://blogforever.eu/wp-content/uploads/2010/11/BlogForever-presentation-vbanos-2010-11-30.ppt

Beagrie, N. (2005). Plenty of room at the bottom? Personal digital libraries and collections. *D-Lib Magazine 11*(6). doi:10.1045/june2005-beagrie

CEDARS Project Team. (2001). Guide to collections management. Retrieved from http://www.ukoln.ac.uk/services/elib/projects/cedars/

Cox, R.J. (2009). *Personal archives and a new archival calling: Readings, reflections and ruminations*. Duluth, MN: Litwin Books.

Coyne, M., Duce, D., Hopgood, B., Mallen, G. & Stapleton, M. (2007). *Significant properties of vector images*. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf

Coyne, M. & Stapleton, M. (2008). The significant properties of moving images. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/preservation/spmovimages_report.pdf

Cunningham, A. (1994). The archival management of personal records in electronic form: Some suggestions. *Archives and Manuscripts 22*(1), 94-105.

Davis, R.M. (2009). Practical blog preservation. Paper presented at the Institutional Web Management Workshop 2009. Colchester, UK.

Entlich, R. (2004). Blog today, gone tomorrow? Preservation of Weblogs. *RLG DigiNews 8*(4). Retrieved from http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file3518.html#article3

Halbwachs, M. (1992). *On collective memory*. Chicago: University of Chicago Press.

Hank, C., Choemprayong, S. & Sheble, L. (2007). Blogger perceptions on digital preservation. *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries. Vancouver, Canada*. New York, NY: ACM Press.

Hedstrom, M. & Lee, C. (2002). Significant properties of digital objects: definitions, applications, implications. *Proceedings of the DLM-Forum 2002: @ccess and Preservation of Electronic Information – Best Practices and Solutions*. Luxembourg: Office for Official Publications of the European Communities.

Herring, S., Scheidt, L., Wright, E. & Bonus, S. (2005). Weblogs as a bridging genre. *Information Technology & People 18*(2), 142-171. doi:10.1108/09593840510601513

Hobbs, C. (2001). The character of personal archives: Reflections on the value of records of individuals. *Archivaria 1*(52), 126-135.

Kaye, J.J., Vertesi, J., Avery, S., Dafoe, D., David, S., Onaga, L., Rosero, I. & Pinch, T. (2006). To have and to hold: Exploring the personal archive. *Proceedings of the SIGCHI conference on human factors in computing systems*. New York: ACM Press.

Kelly, B. (2010). Moving from personal to organisational use of the social web. Paper presented at Online Information 2010, London, UK.

Kelly, B. & Guy, M. (2010). Approaches to archiving professional blogs hosted in the cloud. Paper presented at iPRES 2010: the 7th International Conference on Preservation of Digital Objects, Vienna, Austria.

Kirk, D. & Sellen, A. (2008). On human remains: Excavating the home archive. Microsoft Technical Report. Retrieved from http://research.microsoft.com/apps/pubs/default.aspx?id=70595

Knight, G. (2010a). Significant properties testing report: Electronic mail. Retrieved from http://www.significantproperties.org.uk/email-testingreport.pdf

Knight, G. (2010b). Significant properties testing report: Audio recordings. Retrieved from http://www.significantproperties.org.uk/audio-testingreport.pdf

Lee, C. (2011). Collecting the externalized me: Appraisal of materials in the social web. In C. Lee (Ed.), *I, Digital: Personal Collections in the Digital Era.* Chicago, IL: Society of American Archivists.

McKemmish, S. (1996). Evidence of me. *Archives and Manuscripts 24*(1), 28-45.

Marshall, C., Bly, S. & Brun-Cottan, F. (2006). The long term fate of our digital belongings: Toward a service model for personal archives. *Proceedings of IS&T Archiving 2006.* Society for Imaging Science and Technology, Springfield, VA.

Montague, L. (2010). Significant properties testing report: Raster images. Retrieved from http://www.significantproperties.org.uk/rasterimages-testingreport.pdf

Nesmith, T. (1982). Archives from the bottom up: Social history and archival Scholarship. *Archivaria, 14,* 5-26.

O'Sullivan, C. (2005). Diaries, on-line diaries, and the future loss to archives; or blogs and the blogging bloggers who blog them. *American Archivist, 68*(1), 53-73.

Paquet, L. (2000). Appraisal, acquisition and control of personal electronic records: From myth to reality. *Archives and Manuscripts, 28*(2), 71-91.

Paulus, M.J. (2007). Blogging for the record: A study of blogging from an archival point of view. *Journal of Archival Organization, 4*(3), 31-41. doi:10.1300/J201v04n03_03

Pennock, M. & Davis, R.M. (2009). ArchivePress: A really simple solution to archiving blog content. Paper presented at iPRES 2009: The Sixth International Conference on Preservation of Digital Objects. San Francisco, CA.

Reich, V.A. (2009). Preservation tools for blogs. Paper presented at the Future of Today's Legal Scholarship Conference 2009. Washington D.C.

Sax, L.J., Gilmartin, S.K. & Bryant, A.N. (2003). Assessing response rates and nonresponse bias in Web and paper surveys. *Research in Higher Education, 44*(4), 409-432. doi:10.1023/A:1024232915870

Summers, A. & John, J.L. (2001). The W.D. Hamilton archive at the British Library. *Ethology Ecology and Evolution, 13*(4), 373-384. doi:10.1080/08927014.2001.9522768

Technorati. (2011). State of the Blogosphere 2011. Retrieved from http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction/

Thomas, S. & Martin, J. (2006). Using the papers of contemporary British politicians as a test bed for the preservation of digital personal archives. *Journal of the Society of Archivists, 27*(1), 29-56. [doi:10.1080/00039810600691254](doi:10.1080/00039810600691254)

Whittaker, S. & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction, 8*(2), 150-170. [doi:10.1145/376929.376932](doi:10.1145/376929.376932)

Williams, P., Dean, K., Rowlands, I. & John, J.L. (2008). Digital lives: Report of interviews with the creators of personal digital collections. *Ariadne, 55*. Retrieved from http://www.ariadne.ac.uk/issue55/williams-et-al/

Williams, P., John, J.L. & Rowland, I. (2009). The personal curation of digital objects: A lifecycle approach. *Aslib Proceedings: New Information Perspectives 61*(4), 340-363.

Wilson, A. (2007). Significant properties report. InSPECT Work Package 2.2. Retrieved from www.significantproperties.org.uk/wp22_significant_properties.pdf

Wunsch-Vincent, S. & Vickery, G. (2007). Participative Web and user-created content: Web 2.0, wikis and social networking, 124. Working Party on the Information Economy. OECD/OCDE. Retrieved from http://www.oecd.org/document/40/0,3746,en_2649_34223_39428648_1_1_1_1,00.html