Youth Deleted: Saving Young People's Histories after Social Media Collapse

Anthony Cocciolo ~ acocciol@pratt.edu
Pratt Institute School of Information and Library Science

ABSTRACT

After twenty years of loosing personal digital files, I wonder if today's youth's digital shoeboxes of memories will be even thinner than my own. To test this notion, the following research question is posed: *When social media collapse, are youth disproportionately at risk of loosing their digital contributions?* To study this, the age demographics of failing or failed social media will be analyzed. The list of failed or failing social media is provided by the Archive Team's "deathwatch," which is a group of "rogue archivists" who save web content in danger of disappearing. Results confirm that when social media collapse, youth are disproportionately at risk of loosing their digital contributions because young people disproportionately use the sites that fail. Personal digital archiving outreach efforts need to continue working to educate young people about the risks of loosing digital content and how to avoid such loss.

Introduction with a Personal Story

I, like so many others, have experienced loss of digital files throughout my life. The first loss was the digital files from my teenage years (1993-1998). When I was thirteen years old, I started a software development company where I made shareware and sold it over America Online and the early Internet. When I went to college and needed to free-up my time, I closed down the company, and left the company files on an old Macintosh at my parents' house. One day when I came home, my parents told me the Macintosh was given away to charity since it was never used. I felt a pang of loss for those records, which included all the source code, compiled software, as well as customer records, which were fortunately on an encrypted partition. Today, I would be interested in viewing my early entrepreneurial efforts; however, very little documentary evidence remains of the company (which I called "Dead Broke Software"), except for some scattered references on websites.

The second wave of loss happened in my late teens and early twenties (1999 to 2007), and during the period when I like most others adopted digital cameras. I had just broke-up with a boyfriend, who had all of our shared photos on his hard drive from the last eight years. At the time, I did not think to get a copy of the contents of his hard drive, always thinking that I could possibly get a copy at a later point. One day I read a Facebook post from him that said his car was broken into and his computer stolen. His material loss was my second major digital loss.

As I move through my thirties, I am certainly more careful with my digital files, even going as far to leave backup copies in different geographic locations. However, as an individual with very little documentary evidence from his youth, I do wonder if our

digital infrastructures have improved such that today's youth will have records from their youth, and will not have to rely on memory alone as I have.

I began to doubt this as I browsed the archive of collapsed social media companies put together by the Archive Team. Archive Team "is a loose collective of rogue archivists, programmers, writers and loudmouths dedicated to saving our digital heritage" by keeping abreast of "shutdowns, shutoffs, mergers, and plain old deletions - and done our best to save the history before it's lost forever."[1] The Archive Team—led by Jason Scott—archives websites using a decentralized approach. Any Internet user can download the "Archive Team Warrior," which will download portions of websites using the open-source program "wget" from failing companies and transmit the results to a central repository. The portions are reassembled at the central repository and have resulted in saving once popular websites such as Geocities, which was shutdown by Yahoo in 2010.[2]

The "warrior" is a Linux installation that runs in the virtual machine VirtualBox.[3] Figure 1 shows the Virtual Machine running with the web browser indicating the status of the download job. At the time of writing, the Archive Team was web archiving social media companies Xanga, Posterous Spaces, and Formspring, which are planning to shutdown or have already shutdown.
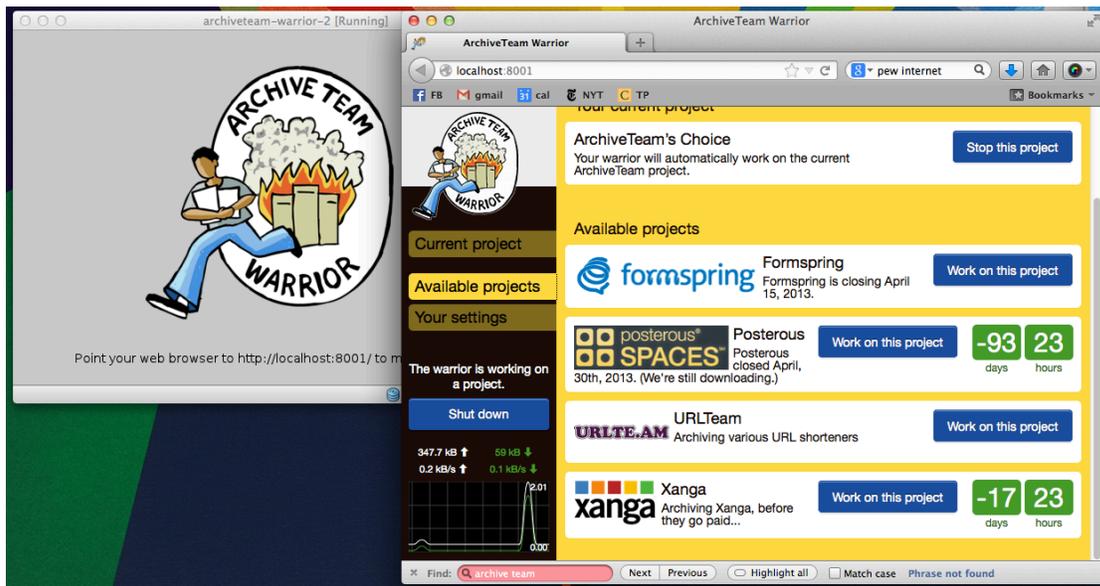

Figure 1: Archive Team Virtual Linux instance for archiving failed or failing websites. Logos copyright respective owners.

In browsing the web archives assembled by the Archive Team, what becomes apparent is how extensively it collects youth culture. For example, browsing the Xanga archive— which the Archive Team is hurrying to archive out of fear of imminent failure—contains profile pages and photos of over 820,000 individuals from the once very popular social network.[4] These profile pages illustrate individuals' attempt to communicate their identity to a larger world, often highlighting their likes, inspirations and ambitions.

Browsing these archives contained on collapsed social media raises a number of questions: Are social media for youth prone to quick fame and equally quick failure? Is it ethical for social media companies to assemble so much youth material with little chance of preserving it for the long-term (other than through the work of rogue archivists)? Will today's youth's digital shoeboxes of memories be even thinner than my own?

To test this notion, the following research question is posed: *When social media collapse, are youth disproportionately at risk of loosing their digital contributions?*

To study this, the Archive Team's "Deathwatch"—which is a listing of sites that have closed down or may close down—will be compared in relation to the demographic groups that use those sites. The method will uncover if youth are particularly at risk, or if their risk is no different than the average Internet user.

However, before the results will be revealed, literature related to efforts to enhance personal digital archiving, as well as efforts to web archive, will be discussed. After which, the study methodology and findings will be discussed.

Literature Review

Social Media and Youth Cultures

Teen use of social media has grown over the last six years. A report from the Pew Research Center found that "Teens are sharing more information about themselves on social media sites than they did in the past" as evidenced in their dataset from 2006 to 2012 (Madden et al., 2013). For example, 91% of teen social media users post a photo of themselves in 2012, up from 79% in 2006 (Madden et al., 2013). Teens use social media primarily to maintain existing social relationships from face-to-face settings, even those relationships that might be weak in nature. The Pew study found that 98% of teenagers use social network site to connect with people they know from school, as compared to only 33% who are friends with people they have never met in person (Madden et al., 2013). Thus, social media scholars such as boyd have determined that "the value of social media rests in their ability to strengthen connections" and youth use social media primarily for "socializing with those they knew from school, worship centers, summer camps, and other activities" (boyd, 2010, p. 113; p. 89).

A notable statistic from the Pew Report is that 60% of teen Facebook users keep their profiles private (Madden et al., 2013). This is a relevant statistic because if a profile is private it cannot be archived through web archiving efforts such as the Archive Team's or the Internet Archive's, since these services depend upon open web content.

The Pew study also found that youth receive pleasure from social media, nothing that "teens are considerably more likely to report positive experiences than negative ones" in using social media and that "52% of online teens say they have had an experience online that made them feel good about themselves" (Madden et al, 2013, p. 2). Despite the pleasure youth receive from social media, some critics may contend that youth—as they

grow to become adults—are embarrassed of their early uses of social media and would not care if such content were eventually deleted. However, the Pew Report indicates that this may not be the case. For example, from the teens surveyed, only 19% reported that they had ever posted an update, comment, photo, or video that they later regretted sharing (Madden et al., 2013). This may indicate that such content posted to social media sites may not be such an embarrassment later in life, and could rather act as a virtual scrapbook of adolescent life and a source of future pleasure.

As youth media scholars such as Ito (2010) have pointed out, social media provide youth with venues to "negotiate issues of identity and belonging within peer cultures" (Ito, 2010, p. 9). She further observes that "just as in the case of youth, who are always on the verge of growing older, media are constantly undergoing a process of aging and identity reformulation in which there is a generation of the new ready to replace the old" (p. 10). Thus, social media come and go, just as youth transition into adults. This however does not mean that youth do not want to refer back to those earlier interactions at some point later in life.

Personal Digital Archiving

Personal digital archiving (PDA) is the practice and study of curating and maintaining personal digital material over the long-term. Marshall (2008) points to a number of challenges to personal digital archiving. One of the greatest challenges with PDA is that "digital loss has a tendency to be an all-or-nothing proposition. People don't lose just a few of the baby pictures of their first child; they lose ALL of them." She believes that if users become active stewards and curators of their own material, rather than practicing benign neglect that is common practice with physical material, then individuals will have a better chance of long-term preservation of their own materials. She also notes that "it is more important to know what we have and where we've put it than it is to centralize all of our stuff into a single repository," which is a fairly common practice. For example, some users email themselves copies of important documents as attachments, effectively treating their email account as their personal digital archive.

Sinn (2011) studied user communications on commercial media providers—particularly email services and blogs—and found that the "general public values personal records in email and blogs" and that the "emotional attachment and preservation attitudes toward the contents in e-mail and blogs were similar to personal papers" (p. 327). Despite user interest in the preservation of their materials, commercial providers take "no responsibility for the inaccessibility or failure to preserve contents" (p. 320). The library and archives community has made calls to the public to actively work to preserve their own person assets, such as through the work of the National Digital Information Infrastructure and Preservation Program (NDIIPP). A notable outreach effort from this initiative is the informational website digitalpreservation.gov/you. Other important efforts include the Internet Archive (IA), which captures the Internet broadly. However, IA does not perform deep archiving, or archiving many page levels below the homepage, resulting in broad yet incomplete archives (Phillips, 2005). Despite calls for attention to this issue from the library and archives community, many contributors to social media

sites are at risk of loosing digital content if their social media site of choice collapses. This is especially true if they don't curate backups of their social media contributions, or have their profiles marked as private, which prevents web archivists like the Internet Archive or the Archive Team from copying the contents.

Lee (2011) argues that many of the traces individuals will leave being are created and distributed through the web. For such content to be available for future use, he observes that web content must be "(1) continuously maintained by the party who is hosting it, (2) preserved by a distributed set of individuals involved in the activities that leave the traces, or (3) harvested by someone with an interest in collection building (e.g., amateur enthusiast, interested scholar, archivist, or librarian)" (p. 203). Since social media companies are startup enterprises that have higher potential of failure than more established institutions (e.g., government, universities), the third option—or web harvesting or web archiving—is a promising solution for long-term preservation of web content.

Web Archiving

The purpose of web archiving is "to collect ephemeral Web content for use by current and future generations of users" (Day, 2006). Reasons often lobbied against archiving web content is that it is not good enough to preserve, it is a self-preserving medium not in need of archiving and that it is an impossible task (Masanès, 2006, "Issues and Methods").

Some of the methods used to archive web content is client-side archiving, server-side archiving, and non-web archiving (Masanès, 2006, "Issues and Methods"). Client-side archiving is the most popular form of web archiving and is used by the Internet Archive's WayBackMachine and by the Archive Team's warrior. In this approach, web crawlers act like normal web users and "start from seed pages, parse them, extract links and fetch the linked document," and re-iterate (Masanès, 2006, "Issues and Methods," p. 23). This method works well for simpler web-pages, but could encounter difficulty when encountering webpages that exchange content in-between webpage loads, which is popularly known as the Asynchronous JavaScript and XML (AJAX) approach to creating web interfaces (Masanès, 2006, "Selection"). Many social media sites make extensive use of this approach, and without special provisions for web archiving, could make retrieving this content challenging. This could be overcome, but may require manual intervention by a skilled web archivist. This approach is also challenged when attempting to download large collections of webpages, which could take a long time to complete download. For example, Masanès (2006) notes that a website with 100,000 pages would take approximately three days to download ("Issues and Methods").

The Archive Team's approach to web archiving—where many distributed users client-side archive websites—is a particularly compelling approach because it greatly enhances the speed in which websites can be downloaded. Many web hosts, if they notice one host is downloading too much content too quickly, will lockout that host or slow-down the rate at which it can download content. Thus, many users who perform client-side web

archiving usually implement a delay between page downloads to prevent being locked-out or overly burdening the webserver from which they are downloading (Masanès, 2006, "Issues and Methods"). Because the Archive Team's approach spreads the downloading across many clients, the delay between page loads becomes more negligible in the aggregate, leading to faster overall web archiving than if using only one client. This contrasts with how most cultural heritage institutions web archive, which is through one or several archiving clients, but usually not through many clients.

Some of the limitations of client-side web archiving are overcome by server-side web archiving, where files are copied directly form the server in conjunction with the site owner's cooperation. This method is currently being used by the Library of Congress to create an archive of Twitter (Osterberg, 2013). The limitation of this approach is re-creating the webpages so that they are authentic to what the user would have experienced, and the extensive effort required negotiating the transfer of data (Masanès, 2006, "Issues and Methods").

Perhaps the simplest form of web archiving is to create non-web archives, where web content is printed out or converted to a format like Adobe Acrobat PDF and stored using something other than the web (e.g., file folders, directories on a computer) (Masanès, 2006, "Issues and Methods"). Although this method has some appeal because of its simplicity, it loses the context in which users experienced the content and the way it was navigated using hyperlinks. It also could loose some of the graphical look and feel of the webpage, which is readily evident when most webpages are saved as PDFs or printed-out.

Study Overview

This study will address the research question: *When social media collapse, are youth disproportionately at risk of loosing their digital contributions?*

To study this, the list of social media sites from the Archive Team's "deathwatch" will be used as the list of failing social media sites,[5] and the demographics for those companies will be retrieved from Quantcast.

The Archive Team's "deathwatch" includes a variety of websites that are in endanger of disappearing altogether as well as websites that may render large amounts of content inaccessible because of takeover by new management. Members of the Archive Team follow industry blogs (e.g., Techcrunch.com) and Twitter feeds from those working in the technology industry to ascertain how imminent the content loss may be. For example, TechCrunch noted on June 4th, 2013, that the entire staff of OMGPOP were laid-off after the buyout by Zygna, which resulted in OMGPOP.com being put on the "deathwatch."[6] At time of writing, that particular site will be shutdown on September 13, 2013,[7] which illustrates the relative speed in which digital content can disappear from the Internet.

Websites from the "deathwatch" that are included in this analysis are social media sites, which are websites that are made-up in large part by user-contributions. For example,

blogging, life-streaming, and photo sharing sites are included in the analysis. Sites that entered the deathwatch in years 2012 and 2013 are included in the analysis. Older websites that have shutdown (e.g., GeoCities.com, which was shutdown in 2010) have been excluded because the contemporary statistics from such sites do not reflect the statistics when the sites were operational.

Included on the "deathwatch" are social media sites that are not necessarily going out of business, but rather sites that have deleted or made inaccessible large amounts of user-generated content. One particular company that will be included with the list of failing social media is MySpace.com, which made inaccessible all of its users blogs, among other content (Williams, 2013). Since the Archive Team was given no notice that these blogs would be made inaccessible, they did not archive them. The Archive Team notes that MySpace has a "penchant for deleting all of your shit with no warning," and thus has declared MySpace.com "endangered."[8]

Quantcast provides demographic information, like age, gender, and income, "for over 100 million websites."[9] One way that it makes demographics particularly useful is by generating statistics for "this site vs. total Internet" for a variety of demographic categories. For example, 18-24 years olds are 2.15 times more represented on MySpace than the Internet in general.[10] They call this statistic "the multiplier." This statistic could also be read as 27% of MySpace users are 18-24 year olds, where only 12% of the general Internet users are 18-24 years olds. If a multiplier were 1.0 for the 18-24 age category, this would mean that the site's audience for 18-24 year olds were represented proportionate to total 18-24 year olds online. Similarly, if the multiplier were less than 1.0, 18-24 year olds would be underrepresented, and greater than 1.0 would be over-represented. For example, the multiplier for 18-24 year olds on aarp.org is 0.49, meaning they are underrepresented compared to the Internet in general. Relatedly, the multiplier for 65+ year olds on aarp.org is 3.66, indicating they are overrepresented on that site in comparison to the Internet in general. This of course makes sense since the AARP provides services to senior citizens in the United States. The statistics used were generated by Quantcast between February 2013 and July 2013. To illustrate trends, the mean and standard deviations of this multiplier will be computed for each social media site and for each age group.

The following social media sites were included on the Archive Team's "Deathwatch," however, did not have Quantcast demographic data and thus could not be included in this analysis: Formspring.me, Jux, 1Up, and Shoutwiki.

Results

Table 1 reveals that young people are overrepresented users of social media sites that are failing, have closed, or are endangered. In particular, youth under 18, 18-24 year olds, and 25-34 year olds are over represented on these sites. The average "multiplier" is 1.10, 1.54, and 1.33, respectively, which indicates these younger age groups are over represented users of these failing websites. Relatedly, 35-44 year olds, 45-54 year olds,

55-64 year olds, and 65+ year old users are under represented on failing social media sites (with average "multipliers" of 0.92, 0.71, 0.53, and 0.43, respectively).

Table 1. Age demographics for failing, endangered or closed social media companies using Quantcast's "multiplier," with Mean and Standard Deviation.

Age/Site	< 18	18-24	25-34	35-44	45-54	55-64	65+
MySpace *	1.73	2.15	1.02	0.6	0.49	0.29	0.19
Xanga	0.6	1.84	1.68	0.79	0.81	0.52	0.5
Posterous	0.54	1.32	1.43	1.1	1	0.7	0.69
Piczo.com *	2.87	1.17	0.64	0.54	0.42	0.29	0.26
Upcoming.org	0.59	0.55	1.35	1.34	1	1.16	0.82
Friendster *	0.6	1.48	2.02	1.12	0.52	0.41	0.23
OMGPOP	0.95	2.32	1.33	0.71	0.66	0.36	0.38
Ning.com	1.29	1.23	1.1	1.15	0.72	0.55	0.35
Tunewiki.com	0.71	1.82	1.43	0.93	0.8	0.53	0.47
Mean	**1.10**	**1.54**	**1.33**	**0.92**	**0.71**	**0.53**	**0.43**
SD	**0.78**	**0.55**	**0.39**	**0.28**	**0.21**	**0.27**	**0.21**

- Quantcast has estimated the demographics for these sites.

This trend where young people are over-represented users of failing social media sites is not true for all sites analyzed. For example, users of upcoming.org (a social media site that was shutdown by Yahoo) were over-represented with users from the 55-65 year old age group. However, in general, failed or failing social media generally have younger user groups.

Discussion and Conclusion

The answer to our research question is yes: *When social media collapse, youth **are** disproportionately at risk of loosing their digital contributions*, because young people disproportionately use the sites that fail.

Media critics such as Rushkoff have pointed to the ways in which media and advertising companies exploit youth for commercial gain. In speaking of Facebook, he (2013) writes that:

> The true end users of Facebook are the marketers who want to reach and influence us. They are Facebook's paying customers; we are the product. And we are its workers. The countless hours that we—and the young, particularly—spend on our profiles are the unpaid labor on which Facebook justifies its stock valuation.

His argument regarding Facebook—which could be extended more generally to social media—is that user contributions provide data that can be mined for targeting of advertising and the content itself is of no particular value to the social media company. Using Rushkoff's perspective, it is not surprising then that social media companies delete or render inaccessible user-contributed content once that content is unable to yield advertising revenues. As with earlier forms of media such as television, social media companies continue to follow a path of using youth for commercial gain where the benefit to the youth is only in the short-term, and a possible long-term consequence include the loss of digital contributions.

Youth are over represented users of failing social media, which indicates that youth are in particular danger of loosing their contributions when social media fails. This is particularly problematic when considering that youth have positive experiences on social media, often don't regret their contributions, and do value their online contributions such as blogs, which could make such sites sources of future pleasure and documentation of one's youth. However, this content is often deleted or rendered inaccessible once it is no longer financial rewarding for social media companies to maintain it.

Personal digital content is prone to loss, especially when social media companies provide an illusory sense of permanence that disproportionately affect young people when they fail. Personal digital archiving outreach efforts, particularly in informal learning contexts, need to continue working to educate young people about the risks of loosing digital content and how to avoid such loss. Through education work, young people have a chance of preserving their digital memories.

About the Author

Anthony Cocciolo is an Assistant Professor at Pratt Institute School of Information and Library Science in New York City, where his teaching and research are in the areas of digital archives, moving image and sound archiving, digital libraries, and educational technology. He completed his doctorate from Teachers College, Columbia University in the Communication, Computing, Technology in Education program, and undergraduate degree in Computer Science from the University of California, Riverside.

Notes

[1] http://www.archiveteam.org/index.php?title=Main_Page
[2] http://archiveteam.org/index.php?title=Geocities
[3] http://warrior.archiveteam.org
[4] http://www.archiveteam.org/index.php?title=Xanga
[5] http://archiveteam.org/?title=Deathwatch
[6] http://techcrunch.com/2013/06/04/zynga-shuts-down-omgpop-one-year-after-acquiring-it-for-200m/
[7] http://techcrunch.com/2013/08/05/omgameoverpop/
[8] http://archiveteam.org/index.php?title=Myspace
[9] http://www.quantcast.com/

[10] https://www.quantcast.com/myspace.com#!demo&anchor=age-gender-container

References

danah boyd, 2010. "Friendship," In: M. Ito et al. (editors), *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media.* Cambridge, MA: MIT Press, at
http://mitpress.mit.edu/sites/default/files/titles/free_download/9780262013369%20_Hanging_Out.pdf, accessed 11 August 2013.

Michael Day, 2006. "The Long-term Preservation of Web Content," In: J. Masanès (editor), *Web Archiving.* Berlin: Springer, pp. 177-199.

Mizuko Ito, 2010. *Introduction,* In: M. Ito et al. (editors), *Hanging Out, Messing Around, and Geeking Out: Kids Living and Learning with New Media.* Cambridge, MA: MIT Press, at
http://mitpress.mit.edu/sites/default/files/titles/free_download/9780262013369%20_Hanging_Out.pdf, accessed 11 August 2013.

Christopher A. Lee, 2011. "Collecting the Externalized Me: Appraisal of Materials in the Social Web," In: C. A. Lee (editor). *I, Digital: Personal Collections in the Digital Era.* Chicago, IL: Society of American Archivists, pp. 202-238.

Mary Madden, Amanda Lenhart, Sandra Cortesi, Urs Gasser, Maeve Duggan, Aaron Smith, Meredith Beaton, 2013. *Teens, Social Media, and Privacy*. Washington, D.C., Pew Research Center's Internet & American Life Project, at
http://pewinternet.org/Reports/2013/Teens-Social-Media-And-Privacy.aspx, accessed 11 August 2013.

Catherine C. Marshall, 2008. "Rethinking Personal Digital Archiving Part 2: Implications for Services, Applications, and Institutions," *D-Lib Magazine,* volume 14, number 3/4 (March/April), at http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html, accessed 11 August 2013.

Julien Masanès, 2006. "Web Archiving: Issues and Methods," In: J. Masanès (editor), *Web Archiving.* Berlin: Springer, pp. 1-53.

Julien Masanès, 2006. "Selection of Web Archives," In: J. Masanès (editor), *Web Archiving.* Berlin: Springer, pp. 71-91.

Gayle Osterberg, 2013. "Update on the Twitter Archive at the Library of Congress," *Library of Congress Blog*, January 4, 2013, at http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/, accessed 11 August 2013.

Margaret E. Phillips, 2005. "What Should We Preserve? The Question for Heritage Libraries in a Digital World," *Library Trends*, volume 54, number 1, pp. 57-71.

Douglas Rushkoff, 2013. "Why I'm quitting Facebook," *CNN*, at
http://www.cnn.com/2013/02/25/opinion/rushkoff-why-im-quitting-facebook, accessed
11 August 2013.

Donghee Sinn, Sue Yeon Syn, and Sung-Min Kim, 2011. "Personal records on the web:
Who's in charge of archiving, Hotmail or archivists?," *Library & Information Science
Research,* volume 33, pp. 320-333.

Rhiannon Williams, 2013. "MySpace users threaten to sue after years of blogs deleted,"
The Telegraph, at http://www.telegraph.co.uk/technology/10173232/MySpace-users-
threaten-to-sue-after-years-of-blogs-deleted.html, accessed 11 August 2013.