digital preservation europe

Considerations for the Preservation of Blogs

Blogs, it seems, are everywhere these days, but what about the next day (and the next and the next …). Opinions vary on whether or not blogs merit preservation beyond the actions of a blog's respective authors. This briefing paper does not contribute to that dialogue. Rather, it provides an overview of issues to be considered by organizations planning blog preservation programs. Blogs are the product of a network of players, including blog authors, service providers, and readers. Discussed here are some key attributes of blogs, and the characteristics and behaviors of these players, which may impact preservation activities.

Introduction

Calls in the literature have advocated that blogs, as potentially valuable additions to the human record, are worthy of stewardship and long-term preservation. Pandora is one example of an organization currently collecting blogs as part of their preservation program (1). Blogs are a ubiquitous component of online life, having emerged in recent years as a pervasive, interactive medium for communication and information dissemination. The extent of the blogosphere, the diverse, heterogeneous aggregation of blog content, is immense (2-3). It is comprised of networks of co-producers made up of blog authors, or bloggers, readers, and service providers (4), reflecting an assortment of associated behavioural, technical, social and legal issues. This briefing paper provides a summary of the key attributes and characteristics of blogs and bloggers which may impact preservation.

Blog Attributes

Blogs come in all shapes and sizes. While there is considerable diversity in blogs' composition, subject, intent, layout and audience, there are some common hallmarks (1-3). Typically, the content of blog posts includes text, images, audio and video, and posts are displayed in reverse chronological order. Content may be original to the blog or the work of others, and may be imported, embedded, or made available through external links. Most blogs are interactive, allowing readers to leave comments, add tags or perform other actions. This two-way communication supports multiple interaction scenarios depending on the number of contributors, size of audience, topical treatment, and availability. Blogs are dynamic and changeable, updated or added to iteratively. This usually happens on a regular basis, though may also be intermittent, depending on the habits of individual bloggers. Blogs are commonly characterized as ephemeral, though the lifespan of blogs varies, from instances that are active – added to, modified, and maintained – for mere days or weeks to persistent instances that have remained active for a number of years.

Blogger Behaviours

Just as blogs are diverse, so are the publishing behaviours of bloggers (1-3). Bloggers publish through a wide selection of blog service providers and host sites. Blogs may be singularly or collaboratively authored. Bloggers self-identify. As such, they may choose to blog under their real names or use a pseudonym. While most blogs are characterized as open, bloggers may limit access to all or parts of their blog. Bloggers may choose among several techniques for composition, including a blog's internal editor feature, or methods external to the blog, such as word processing programs, email composers, or desk-top plug-ins. Bloggers may easily update, modify and alter their blogs, both in terms of the look, or layout of the blog, but also in terms of content. Blogs, including posts and comments, may be intentionally edited or deleted. Such activity supports the notion of the ephemeral nature of blogs, not only in the publishing of new content, but in altering or loss of previously published content. Further, unintentional or accidental deletions or loss may occur. In a well-publicized example, the Google Blog was "accidentally deleted," and then recovered once the error was detected (4).

Further information and resources:

(1) Pandora, http://pandora.nla.gov.au/

(2) Technorati. "State of the Blogosphere/2008." http://technorati.com/blogging/state-of-the-blogosphere/

(3) Amanda Lenhart and Susannah Fox, Bloggers: A Portrait of the Internet's New Storytellers. Pew Internet and American Life Project (July 19, 2006). http://www.pewinternet.org/pdfs/PIP%20Bloggers%20Report%20July%2019%202006.pdf

(4) Vivian Serfaty, The Mirror and the Veil: An Overview of American Online Diaries and Blogs. Amsterdam and New-York: Rodopi, 2005.

(5) The Official Google Blog. "And We're Back" (March 27, 2006). http://googleblog.blogspot.com/2006/03/and-were-back.html

(6) Priscilla Caplan, "The Preservation of Digital Materials," Library Technology Reports 44, no. 2 (February/March 2008): pp. 1-9. https://publications.techsource.ala.org/products/archive.pl?article=2614

(7) E.g., BlogBackupOnline, http://www.blogbackuponline.com/techrigy/; BlogBackupr, https://blogbackupr.com/; Archive-It, http://www.archive-it.org/

Preservation Considerations

These general blogger and blog characteristics lead to several considerations if a blog is to be "preserved." Caplan summarizes the goals of digital preservation as: availability, identity, understandability, fixity, authenticity, viability, and renderability (5). An exhaustive treatment of issues that arise when considering blog preservation in relation to these goals is outside the scope of this briefing paper. Provided here is a sample of only a few issues in relation to the first three goals:

• Availability: Any preservation action requires the simple step of duplication. Permission to duplicate blog content requires negotiating among the rights of multiple co-producers: the blogger, or bloggers in cases of collaboratively-authored blogs; contributing commentators; blog service providers, and other content producers, as in the case of embedded and imported content.

• Identity: A number of factors impede the ability to adequately describe the blog. For example, the use of aliases complicates determinations of authorship, as well as establishing other controls, such as permissions and credibility and authentication. While the use of descriptive tags is a common feature of most blogs, the quality and accuracy of such identifying information varies widely.

• Understandability: Is preservation of content, regardless of context, sufficient? The inability to capture and preserve the design and features of the blog contradicts defining attributes of a blog, including interactivity and periodic, chronological publishing. Most bloggers make use of service providers, and as such, their blogs are subject to the providers' terms of service, and dependent on the continuation of such services for the publication and availability of their blogs

In addition, the diversity of blogs and differing perceptions of quality and value leads to questions of appraisal and selection. Blogs are not created equal. There are differences in audience, readership, domain, credibility, and authority. Should all blogs be preserved? This question is very different from "can all blogs be preserved." For the latter, no. There are simply too many mitigating factors for this to be achieved, including the shear volume of active and inactive blogs, issues of ownership and copyright, and issues of access and availability. In this light, the former question may be rephrased, "which blogs should be preserved." Obviously, an individual blogger may select to take actions to preserve, in the most basic sense, their respective blog. Tools are being made available online to facilitate personal blog preservation (6). For organizations administering blog preservation programs, clear parameters for selection are essential.

Conclusion

The considerations presented here illustrate the range of issues that impact the preservability of blogs, and the critical decisions that are required before preservation programs are implemented. Such decisions will be enhanced through an informed understanding of the characteristics and behaviors of a program's targeted community of blog co-producers, including bloggers, service providers, and readers.

Carolyn Hank, Laura Sheble and Songphan Choemprayong, *School of Information and Library Science, University of North Carolina at Chapel Hill*
hcarolyn@email.unc.edu