Pamela J. Smith

Handling New Media

December 14, 2004

Final Paper

Eyebeam.org/reblog

Assessing the Risks of a ReBlogging Web Log

A new web form

The web log or "blog" is a specific form of web page that proliferates throughout the web, allowing users to publish electronic content second by second, day by day. The blog exists in a public space yet there is a sense of immediacy and freedom in a blog post that creates a kind of personal-now-public record. News organizations and individuals who want to constantly compile and refresh content are the most popular users of the form. So popular that the Seattle Post-Intelligencer reported that the most requested online definition this year was "blog," a word that is not even officially printed in the dictionary yet.[1] Software has been developed that structures the newest posts at the top of the page, with the most recent entries refreshing at the top, accumulating a list of date and time-stamped entries. Some packages include tools that make it easy for users to format the text, create hyperlinks, add article summaries, images or audio, and enable comments. Most packages also have a way of archiving older entries. The web log is the new information resource and the new diary.

The form accumulates a unique currency of historical documentation that must be preserved. Yet long-term preservation of web-based material is a formidable challenge with

[1] Tynan, Trudy. Seattle Post-Intelligencer. 12 December 2004. Seattle Post-Intelligencer. 12 December 2004. http://seattlepi.nwsource.com/business/aptech_story.asp?category=1700&slug=Dictionary%20Top%20Words. See wikipedia.com http://en.wikipedia.org/wiki/Blog for a comprehensive definition and history of the web log.

numerous problems, risks and special needs.[2] Various organizations have developed strategies for preserving digital information, yet there is no comprehensive standard. Using Eyebeam's blog-based website www.eyebeam.org/reblog as a case study, I sought to assess a work born and living on the web that is active, interactive and growing. By documenting the history, behavior and content—both the underlying structure and the site as a whole—I attempt to assess the site's special needs and determine the "best" guidelines to preserve such a work.

Eyebeam reBlog

The reBlog site posted at the end of November 2003 after five months of development in Eyebeam's Research and Development Lab led by Jonah Peretti and Michael Frumin. The site was designed by Ann Poochareon and James Daher. The reBlog system is a hacked version of Movable Type 3.11 web log publishing software. ReBlog 1.1 software is publicly available as open source software that anyone can download and install and use for reading and republishing blog content. Essentially the reBlog program republishes packets of syndicate information from multiple websites. This makes it easier for users to browse blog content with the option to read more detail from the original source and/or republish posts. Every month a new reBlogger curates content streaming from RSS (Really Simple Syndication) feeds so that each month the thematic focus of reBlog changes according to the taste of the reBlogger, with an emphasis on new technology, art and politics. The guest reBlogger can edit the content as well, and add comments with each post. The exchange of a program automatically filtering feeds with a human selecting, reformatting and attributing the content adds a level of variability and

[2] Besser, Howard. "Digital Longevity" in *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Ed. Maxine Sitts. National Document Conservation Center. 2000. http://www.gseis.ucla.edu/~howard/Papers/sfs-longevity.html. 3 December 2004.

interactivity to the work that cannot easily be contained. The site depends on the exchange of a community.

Ann Poochareion, designer of the site and guest curator last month, described the process and experience of reBlogging on her personal blog, misery.net, "There's also more incentive for you, as a human reblogger, to know that other people are also checking out your site and taking in your feed that you've filtered as more information for their brain. In the end, what you have is an information portal site, like so many that already exist on the Internet now, but imagine if you too take in the Associated Press feeds or Reuters feeds and start becoming one of the portal sites. Think about this for a bit in your shower or during your morning coffee. Think about how this *could* (and is) changing the way the public receives information about any news, any political situations, any personal journal, any op-eds and debates, any science study and discoveries, any photos from around the world, any video broadcasts…"[3] The site currently serves as a portal for 136 feeds of varying amounts of information, in various forms, and a curator can post as much as he or she likes. Other reBloggers liken the filtering process to DJ-ing, "an art, somewhere between curating and editing"[4] Some reBloggers have reBlogged posts from other reblogs ("re-reBlogging"[5]) and vice versa, some bloggers have reblogged posts from reBlog.

When handling an object that draws from a large network of data and depends on exchanges and links, the archivist must analyze the essential behavior of the work in relation to its creator(s), as well as look at the underlying software and hardware that facilitates this behavior. Where does the provenance of the work begin? What is the relationship between

---

[3] Poochareion, Ann. "so, what the hell is a reblog?"1 December 2004. miserychick.net. December 1, 2004.
[4] Moody, Tom. "Last Day reBlogging - reBlogging Philosophy" 22 September 2004. http://www.eyebeam.org/reblog/archives/2004/09/last_day_reblogging_reblogging_p.html. 10 December 2004.
[5] Shey, Tim. "on reBlogging" 4 October 2004. http://shey.net/reblog/2004/10/ive_just_wrappe.html#more. 10 December 2004.

original content (blogged) versus mediated content (reblogged)? How many links are there? What are the limits of this work? What can be preserved within these boundaries?

Description of the component parts

The major features of a reBlog post are: the title (usually a link to the original source post), content summary (sometimes with additional links by subject), an image, an attribute to the original post (with link), an attribute to the original blog site (with link) and a note designating who reBlogged and when by date and time (with link to reBlog summary). The main page at www.eyebeam.org/reblog/ includes a photo and caption of the guest reBlogger of the Moment and lists the most recent reBlog posts first, accumulating at the bottom of the page a list of recent entries by title. The main page's masthead on the left-hand side of the screen also consists of a running list of reBlog's feeds, an Archive function that organizes posts by month and year, and a search field driven by Google that users so that can search the site's contents.

The latest headlines, with hyperlinks and summaries, are fed into reBlog in the RSS or Atom XML-format, to be read with a feed reader. RSS documents employ a set of tags to describe the major features of the text (such as title, author, link). The reBlog system consists of two main components: reFeed, a web-based RSS aggregator, derived from open source Feed on Feeds[6], and a blog publishing platform based on Movable Type version 3.11. The system itself is publicly available as an open source software package (currently reBlog version 1.1). As a web-based aggregator, software installation is not required, and the interface is available on any computer with web access.[7] ReFeed is a friendly interface that enables the user to view and add RSS feeds, keep track of streaming content, mark feeds as read or unread, with the added

---

[6] Minutillo, Steve. "Welcome to Feed on Feeds." <//minutillo.com/steve/feedonfeeds/> 10 December 2004.
[7] Wikipedia. 13 December 2004. http://en.wikipedia.org/wiki/RSS_%28protocol%29 13 December 2004.

functionality to publish or not publish, and format the feed according to his or her specifications before it's published (including text and image). Users can change the title, primary link or content embedded within the original RSS feed, and add comments and subject tags. The post can also be previewed. These features for republishing and redesigning posts—the curatorial function—is what makes reFeed different from other feed readers. The repost along with the added attribution of the blog source (title, URL, feed of the original source and reBlogger) is output in standard RSS 1.0 format. ReFeed is distributed under general public license (GPL), making it free to all users.[8]

RSS feeds are written in a language based on XML (eXtensible Markup Language). XML is a W3C recommendation for creating special-purpose markup languages; its primary purpose is to facilitate the sharing of structured text and information across the Internet. It is a popular language, and as it's platform-independent, it is relatively immune to changes in technology.[9] RSS is a popular protocol, widely used by numerous news websites and blogs. It is such a common syndication format that it is supported by Movable Type's default templates.

The blog publishing software Movable Type is written in Perl (Practical Extraction and Report Language), an open-source programming language that supports several advanced features and add-ons. Perl is often considered the archetypal scripting language and has been called the "glue that holds the web together," as it is one of the most popular CGI languages. Perl is also free software, available under a combination of the Artistic License and the General Public License. It is available for most operating systems but is particularly prevalent on Unix and Unix-like systems (such as Linux, FreeBSD, and Mac OS X), and is growing in popularity

---

[8] General Public License <http://www.reblog.org/refeed/LICENSE. 10 December 2004.
[9] Wikipedia. 11 December 2004. <http://en.wikipedia.org/wiki/XML> 14 December 2004.

on Microsoft Windows systems.[10] Developer Frumin changed the language of Moveable Type plug-in software to create reBlog's platform. Movable Type allows users to format and save blog entries, post them chronologically with the most recent first, and generates a specific page for each post with a unique URL that is searchable and saved separately as part of the Archive. Movable Type is not an open source since technically Six Apart, Ltd owns it, but it's free to download, modify or create derivative copies for personal, non-commercial use. Licenses (and corresponding pricing schemes) are given according to educational, commercial or not-for-profit use. Movable Type is widespread, with broad platform support for various operating systems and web servers.[11]

The archive

Movable Type software offers default support for archiving by individual post, post category, or by date groups such as monthly or weekly. ReBlog software generates a specific page by individual post, with the date and title in the URL. This is an invaluable feature for the archivist wanting to track and control individual pieces. With each page being searchable, one can look up by title or date for instance, and organize the pieces by subject or by provenance. Also, by handling the work on an item-level basis, beginning with the entries posted each day and gradually connecting the pieces and expanding the network, the archivist could begin understanding the production process. The Archive's contents are listed by month and title in the reBlog masthead along the left-hand side of the screen.

The Archive function saves most of the pieces of each individual static post, including title, content image and attribute, along with hyperlinks, assigning a URL that points to the post's

[10] Wikipedia. 13 December 2004. http://en.wikipedia.org/wiki/Perl 13 December 2004.
[11] movabletype.org: Product Overview. 2001-2004. http://www.moveabletype.org/product_overview.shtml 14 December 2004.

physical place on the server. However, the Archive function technically does not preserve the main page, as it existed at the time, because the reBlog masthead changes each time the site is rebuilt. Thus the picture and attribute of the guest reBlogger may not match the person posted in the archive. Frumin recognizes this as a flaw in Movable Type that should be fixed. He says there is talk that the program will soon be accommodating an author-based Archive that will index and locate developers within the original posts' URL.

The Archive function does not save the functionality of the reBlog in that it fails to include all blog content linked from the reBlog site. The reBlog points to other server directories for its content, resulting in broken links if server-side content, software or hardware disappears on the other server's end. Images are also linked on the server-side as well; this is a major weakness in the reBlog software not only because it becomes difficult to save posts as they existed, but there is vulnerability in the work even as the post exists now—bloggers can easily change the image once it is fed and posted on the site.[12] This leads the Archivist to wonder what the boundaries of the reBlog site are. It is important to determine not only what is original and what is reBlogged, but also determine what is interactive and changeable.

Although in theory most blog software packages include an Archive function it serves as more of a means to organize or file content away and does not authoritatively save and preserve all active data comprising the reBlog site as a whole, including the text, links and images their references.

---

[12] See an example of a post that was later changed by hacker Andrew Baron 28 November 2004. http://www.eyebeam.org/reblog/archives/2004/11/reblogingtxtimg.html 2 December 2004. Baron posted on his own blog the "very easy" steps he took to perform his conceptual switch in "Today I Hacked reBlog: The End of Era for Data Control." 29 November 2004 < 2 December 2004. Noting the difficulty of authenticating web content, Baron writes, "You can see how dynamic reBlog's archival content must be, as people drop, clean, and change their server data, often deleting their own past, directly effecting [sic] Eyebeam's past."

Risk assessment

ReBlog relies on multiple programming languages and the platform Movable Type. The languages reBlog incorporates, Perl and XML, and the RSS protocol, are considered versatile, open source components and widespread amongst users. Perhaps obsolescence is not such an immediate concern. Though also popular and compatible with most operating systems, the platform Movable Type, on the other hand, is open source with restrictions and Six Apart Ltd. technically owns and controls the software under license. Though it is written with an adaptable language, Perl, reBlog is at risk of Six Apart's power to stop Eyebeam from distributing its hacked version. One wonders if Six Apart has the power to control content already posted by its hacked software. Though a possible lawsuit may not affect archived content, it may become difficult to reformat content in the future without Movable Type updates. Also, if Movable Type were to go under, Eyebeam would lose its platform support.

Multiple versions of the reBlog software exist: 0.1, .9, 1.0 and 1.0. All versions are saved on Eyebeam's servers; all versions except for 0.1 were made available online for downloading. As the software continues to change, pages created by the original program are vulnerable to incapability with pages created by future software. It is key to translate components of the software clearly between versions so that data is not changed or lost in the conversion. Also it is important to back-up the original files inside and outside the program—in all its incarnations—on multiple servers. (Frumin already uses CVS software to track and save the files for his programs on the Eyebeam server—this is called versioning software). ReFeed as a reader that also risks becoming incompatible as new versions are developed. The archivist must anticipate how reformatting may affect the understandability and the usability of the work.[13] Frumin also

[13] Besser. "Digital Longevity" in *Handbook for Digital Projects: A Management Tool for Preservation and Access*.

noted how Eyebeam wants to change its design to accommodate the Eyebeam logo on both the main page and individual archived pages. This challenges the authenticity of original reBlogged posts and changes the history of the reBlog site.

Content is also vulnerable to developers outside of Eyebeam. The text is stored and controlled by Eyebeam via reFeed, but the links and linked content are stored by both Eyebeam and the original blog sites, and the images are stored solely by the original blog sites, completely out of Eyebeam's control. ReBlog relies on other sites to archive their content, which is risky considering current blog software is not up to archival standards in authenticating and saving content based on the original post (though it does reference the specific place from which it originated). Tracking images is especially problematic since RSS feeds do not carry the image itself but only its referent. Also since the Archive does not automatically record the functionality of exchanges between multiple servers at the time of original posts, links to original content are in danger of being broken. The site's fragility becomes evident when one plugs its URL into the Internet Archive. Using a form of migration strategy, the Wayback Machine's crawler captured the files from the reBlog site February 5, 2004, along with its Archived pages from January, December and November. Although the text is complete, some of the links are broken and some of the images are missing (including the photograph of the reBlogger of the Moment, Jonah Peretti).[14] This illustrates that the Internet Archive's methodology is not enough. The work is not just a static composite of text and images but depends on server functionality.

The archivist must determine if the limits of the work are physical and/or conceptual—perhaps the limits are all the different ways to use the site. The archivist must decide if she must preserve the content of the work, the data, or simply the "look and feel" or basic behaviors of the

---

[14] www.eyebeam.org/reblog archived. 5 February 2004. http://web.archive.org/web/20040206011521/www.eyebeam.org/reblog 13 December 2004.

work. The archivist must distinguish between the parts that are essential and those that are contextual.

Possible preservation strategies

Jeff Rothenberg begins his report, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*, with the grim declaration, "There is yet no viable long-term strategy to ensure that digital information will be readable in the future."[15] Works or documents born digital are vulnerable to loss via the decay and obsolescence of the physical storage media that holds the data, and become unreadable and inaccessible if the software or hardware needed to interpret the data become incompatible, lost or obsolete. Works that are dynamic, distributed and interactive must also retain their functionality. Rothenberg argues that the functionality, look and feel of the original document is more important to retain than the original medium which is subject to rapid decay.

As the reBlog site is expansive and interactive, linking files on Eyebeam's server with files on the recipient's server, it is crucial for the archivist to determine the boundaries of the site and preserve the essential elements within these limits. Since the reBlog cites content of a fleeting nature, from news sources that are not necessarily preserved elsewhere (with the exception of feeds from such large media institutions like the BBC or the New York Times), I think the historical documentation of art and technology ephemera should be preserved, including all content that links from reBlog. This includes the entry reBlogged on Eyebeam's site, the blogged entry, and the source that initially posted and started the chain of linked information (sometimes the blog and the original source is the same). It is essential to

---

[15] Rothenberg, Jeff. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington: Council on Library and Information Resources, 1999.

continually check and update links. Preserving such relationships requires extensive documentation. Richard Rinehart suggests a notation for media art that acts as a score or "set of instructions that trigger actions or events." This notation would record not only the "location of files and objects, but also the explicit declaration of behaviors, interactions, choices, contingencies and variables." [16]

Using controlled vocabulary, it is essential to bring together documentation generated from the production process as well as actively document the experience of those involved in the project, including developers, guest reBlog curators, individuals who use reFeed to write and republish blogs, and those readers who read reBlog. It would also be useful to visit blog sites of guest reBloggers and record their thoughts about reBlog.[17]

In the emulation approach, Rothenberg suggests that key pieces of the strategy have worked in the past.[18] Assuming migration to be a laborious and fruitless method, he suggests bundling digital documents with their original software, with the added protection from being access by any other inappropriate software, to be run on an emulator. Rothenberg emphasizes that tests have only run the software on known platforms so that longevity is still a question. Like Rinehart, Rothenberg suggests accompanying as much "explanatory documentation" as possible, "including explanations of how to use the encapsulation itself, user documentation, version and configuration information for all software that is to be run under emulation…and the emulator specification itself." Key documentation for the reBlog software includes the reBlog and reFeed websites (including download instructions) and the software read me text files included in the download. Since the reBlog software is available to download, multiple versions

---

[16] Rinehart, Richard. "A System of Formal Notation for Scoring Works of Digital and Variable Media Art." www.bampfa.berkeley.edu/about_bampfa/formalnotation.pdf 13 December 2004.
[17] Beth Rosenberg, Director of Publications for Eyebeam, has already started a compilation of guest reBlogger comments.
[18] Rothenberg. *Avoiding Technological Quicksand*,

of the software are out there. This is another preservation strategy to consider—redundancy. If Eyebeam's servers crashed, Frumin could track down a downloaded copy from someone, somewhere.

Although these are just two strategies gleaned from hypotheses, extensive documentation is necessary for both the physical, structural components of the work along with the possible connections, interactions and choices driving these components. Additional case studies and applying emulation theory to practice are sorely needed in the field.