Case Study: Science, Medicine, Mathematics, and Technology blogs

Background

The mission of the National Digital Stewardship Alliance is to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations. The NDSA's Content Working Group is focusing on investigating guidelines for the selection of significant content, discovery of at-risk digital content or collections, and engaging all stakeholders in the process of acquiring, preserving, and providing access to the content. Working group members are developing case studies and/or models to share broadly with all stakeholders, from content producers to cultural heritage organizations. The goal of this case study is to engage all members of the community in the preservation of content and to encourage the cultivation of relationships that could enable preservation. This case study addresses at-risk science, medicine, mathematics, and technology blog content.

About Science, Medicine, Mathematics, and Technology blogs

Science, medicine, mathematics, and technology blogs include everything from blogs that professionals keep about their personal lives and activities, to open notebook science where practitioners are sharing their daily raw notes, to more reflective and commentary based blogs by individuals, professional organizations, and institutions. Included also are blogs about science, medicine, mathematics, and technology that document the perspectives of the many other individuals and groups who engage in these topics, including the voices of citizen science, healthcare, science policy, and science news.

Historical Value

Blogs on topics in science, medicine, technology, and mathematics have significant research value for historians, sociologists, and anthropologists of the twenty-first century as they:

- Reflect and may influence a wide range of viewpoints, and capture dissenting opinions and new trends not found in other publications
- Serve as a record of who did what and when and will serve as a data mine for understanding trends in scientific practices and attitudes, and how ideas evolve
- Provide insight into the environment of science, medicine, technology, and mathematics
- Show how science, medicine, technology, and mathematics is communicated among collaborators and broadly in the scientific community

Blog content is at risk for loss due to its ephemeral nature. Content may be intentionally or unintentionally deleted by the creators of the blog or the service providers who host them.

Some organizations already engaged in collecting science, medicine, technology, and mathematics blogs include the Bibliothèque Nationale de France, Library of Congress, U.S. National Library of Medicine, National Library of New Zealand, and Smithsonian Institution Archives.

Recognized Opportunities

Social media and blogging platforms don't readily allow via technical infrastructure ways for content creators to easily download and save their own content. Terms of service on social media sites often restrict copying for preservation by libraries and archives, even if content creators are amenable to preservation. NDSA can work to reach out to content creators and publishing platform creators to raise awareness of preservation requirements, and to engage stakeholders in selection of important content for preservation.

Target Audiences

Those who will likely find value in this content include:

- The creators of blogs by practitioners who want long term access to blogs that document their work, thought processes, successes, and failures
- Scholars, including historians, anthropologists, and sociologists who are seeking to understand what it was like in the world of science in the early twenty-first century: what topics were debated, the controversies, what it was like to be a scientist, how scientists worked together, how science and culture intersected, how scientists engaged with the public, etc.
- Future practitioners who want to build upon earlier work
- Professional associations who want to preserve the work, reflections, and communication of ideas of their community
- The creators of the more personal blogs, and their families. Blogs may document personal experiences and reflections on life, and are a means of preserving memory. This content is similar to analog diaries and journals that are passed down in families, and sometimes donated to libraries and archives
- Archivists and librarians whose role it is to preserve the culture of science, medicine, technology, and mathematics

How can we get these stakeholders involved?

- Creators can be encouraged to take steps to make their blogs more preservation friendly
- Creators may have professional obligations to universities or institutions to create and preserve blogs as documentation of their research
- Scholars can share their research based on blogs to convey value in blog preservation
- Professional associations and universities can communicate to their members the value of preserving blogs, and can communicate guidelines and best practices
- Archivists and librarians can be a resource for best practices and guidelines for personal digital archiving

2

- Archivists and librarians can collaborate to preserve and make accessible blog content that fits within their institutional collection development policies.
- Individuals can share personal experiences and value of using blogs to understand the past. These stories will help librarians and archivists persuade communities to take action.

Educating Stakeholders

The NDSA seeks to communicate the need for preservation of blogs on topics in science, medicine, technology, and mathematics to stakeholders through multiple venues:

- Communicate guidelines and best practices to blog creators though personal digital archiving efforts/programs, through professional associations with members who both create blogs and care about future access of blogs related to their community
- Support for librarians/archivists who are seeking a better understanding of the tools, techniques, and standards for Web archiving
- Work with the decision makers of libraries and archives on the value of blog preservation (and web archiving in general)

Obstacles and Risk Factors

Possible risk factors and obstacles in Science, Medicine, Mathematics, and Technology blogs preservation include:

- Scope: We will not be able to capture everything we would like. How do we decide what to preserve? What can we not preserve?
- Rights: If we are preserving the blogs of others, do we need permission to capture? What if there are multiple authors? Do we need permission to make them publicly accessible? ARL Code of Best Practices for fair use might be helpful
- Perceived lack of value: Some bloggers will choose not to take necessary steps to preserve their blog; some bloggers may not want to preserve their blogs. It is very possible that the historical value of a blog will be unknown until much later, long after preservation actions can be taken
- Technical limitations: challenges in preserving social media content with current crawler and access tools, and keeping up with bloggers moving to a new platform or URL.
- Limited resources: Libraries/archives with programs archiving web content outside their domain are best positioned to preserve this content. Are resources sufficient to support this work broadly?
- Archival collection development practices that focus on the acquisition of collections at the end of a donor's career or life
- Funding for digital preservation may not be readily available
- Standards: There are many different and varied digital preservation strategies which make it more difficult to educate stakeholders
- Awareness: Different levels of understanding among scientists about the importance of preservation

Actionable Items

The NDSA and identified stakeholders have an opportunity to address the concerns outlined in this case study. Some ideas that have emerged from discussions within the Content Working Group include:

- Create a registry of science, medicine, technology, and mathematics blogs and identify which ones are currently being preserved. This will allow libraries and archives to avoid or at least know about duplication of effort
- Promote principles of personal digital archiving. Are there clear and simple guidelines and best practices for blog authors to follow if they want to take action to preserve their blogs? (Limitations: Some blog authors will not have the resources, time, expertise, or incentive to act to archive their own content. These blogs would be at a higher risk for loss. Even if a copy of blog is saved on author's own computer, it will likely be a preservation concern later down the road if the creator does not migrate data over time to new media or systematically make copies of the updated content. Public access to blog content preserved by individuals may also be limited)
- Collaborate with the producers of major blogging platforms, like WordPress, Blogger, and Moveable Type, with large numbers of subscribers, to promote best practices for personal blog preservation and potentially match the blog to a host institution for long term preservation. (Limitations: Would miss content not available through major blogging platforms and from those that do not opt-in.) The NDSA Content Working Group Blog Preservation Action Team has already taken steps toward developing a plugin to allow for blog authors to opt-in for preservation
- Create opportunities for bloggers to connect with organizations and institutions that can preserve their content

For more information, to provide feedback, or to get involved, contact ndsa@loc.gov (please include the subject line: Content Case Studies).

4