Bloggers on Digital Preservation: Key Findings and Future Directions

Carolyn Hank
Songphan Choemprayong
Laura Sheble

School of Information and Library Science
University of North Carolina at Chapel Hill
http://persistentblog.wordpress.com

CRADLE, SILS-UNC: 8 Feb 2008
Dynamic, fluid compositions

Networks of service providers and content co-producers
Bloggers

57%

post several posts a week
Bloggers

~5 hours working on blogs a week

(Mode = 1 - 2 hours)
Bloggers

80%

read other blogs daily
Method:

Web Survey
46 questions
Active bloggers

Snowball Technique

223 completed surveys

core group
33
Survey Topics

- Awareness
- Appraisal
- Impact
- Investment
- Demographics
Demographics

- 55% male
- 45% female
- 24% outside the US
- Nearly 2 of 3 graduate or professional degree

Age Distribution:

- <20
- 20s
- 30s
- 40s
- 50s
- >60
43% have blogged for 3 years or more
41% contribute to two or more single-author blogs.

47% contribute to two or more collaborative blogs.

Bloggers
Identity

56%  20%  24%

Real name | Both | Pseudonym
Topical Areas

Professional Technology Personal

Academic News Politics

Hobbies Entertainment Family/friend

Religion Business Health Sport
Preservation

71% agree that their own blogs should be preserved

36% agree that every blog should be preserved
Hosting

16.10% | Other non-commercial server
14.50% | Aggregated blogs
13.50% | Commercial blog provider
16.10% | Server at school
7.30% | Social networking site
3.60% | Server at work
3.60% | Commercial server
1.60% | Own server

Blog service provider 86%
Often use

**blog editor**

Alternative way:

Word processor (desktop and online)
Email composer
Mobile phone/PDA
Desktop plug-in blog editor
Instant messenger
2% require password to view ALL posts

19% require password to view CERTAIN posts
96% edited content after it was posted

EDIT Posts

Spelling & grammatical errors
Rephrasing
Updating links
Erroneous information
DELETE Posts

39%

deleted content
after it was posted

Remove opinions no longer held
Remove info no longer correct
Erroneous information
DELETE Entire Blog

23%

have deleted an entire blog
84% have changed the appearance of their blogs.
67% duplicate some posts
(18% duplicate all)

36% duplicate some comments
(12% duplicate all)
Characteristics

- Unique/innovative content
- Cultural significance of bloggers
- Preservation preference
Characteristics

* Number of readers
* Age of blog
* Size of blog

Unimportant
Selection Criteria

- Context
- Content
- Style
- Format
- Audience
- Author
Selection Criteria

- Blogosphere
  - (all, as many)
- Representation
  - (time/society)
- Future use
- Significance
  - (cultural, social, political)

context
content
style
format
audience
author
Selection Criteria

Subject
(e.g. art & culture, politics & government, science & technology, sociology, education, history)

Attribute
(original, innovative, quality, authoritative)

context

content

style

format

audience

author
Selection Criteria

- Expressive
- Citizen journalism
- Commentary
- Advocacy
- Scholarship
- Utilitarian

- context
- content
- style
- format
- audience
- author
Selection Criteria

- Context
- Content
- Style
- Format
- Audience

Text Photos
Video Audio Other_multimedia
Comments Keywords Links
Imported_images Bookmarks
Imported_tags Feeds
Selection Criteria

- **Popularity**
  - (e.g. readership)

- **Impact**
  - (e.g. reference, citation)

- **Location**
  - (e.g. regional, community)

- **Public discourse**
  - (cultural, social, political)

- **Audience preference**
- **Expert preference**

- **context**
- **content**
- **style**
- **format**
- **audience**
- **author**
Selection Criteria

- Context
- Content
- Style
- Format
- Audience
- Author

Options:
- Professional
- Personal
- Institutional
- Collaborative
- Expert & Notable
- Author preference
Exclusion Criteria

- **context**: ephemeral & insignificant
- **content**: personal & inappropriate
- **style**: hate speech
- **format**: splog, storage concerns
- **audience**: popularity
- **author**: personal
“[S]hout, plead to my boss not to fire me!”

“I would seek back-up files from the sys admin, as I have a hard time believing that a blog would just disappear.”

“[D]rink some coffee and do other hobby.”

“I have no sense of identity invested in my Weblog, so I probably wouldn’t do anything, but simply let it go.”
54% would not pay to preserve their OWN blog.

Annual: $1 – 200
Average: $36.82
**Findings**
- Bloggers are interested
- Save some but not all
- New content added
- Old content altered
- Personal responsibility
- Defining roles of others

**Future**
- Methodology
- Responsibility
- Access scenarios
- Versioning
- Intellectual Property
- Access scenarios
- Process in time
THANK YOU

For Advising…
Paul Jones
And Advice…
Helen Tibbo and Cal Lee
Curation & Archives Research Group