

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

Blogosphere

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

core group

33

223

completed surveys

Web Survey

46 questions

Active bloggers

Snowball Technique

Method...

Survey Topics

Awareness

Appraisal

Impact

Investment

Demographics

Demographics

55% male
45% female

24% outside
the US

Nearly 2 of 3
graduate or
professional
degree

<20 | 20s | **30s** | 40s | 50s | >60



Bloggers

43%

have blogged for 3 years or more

41%

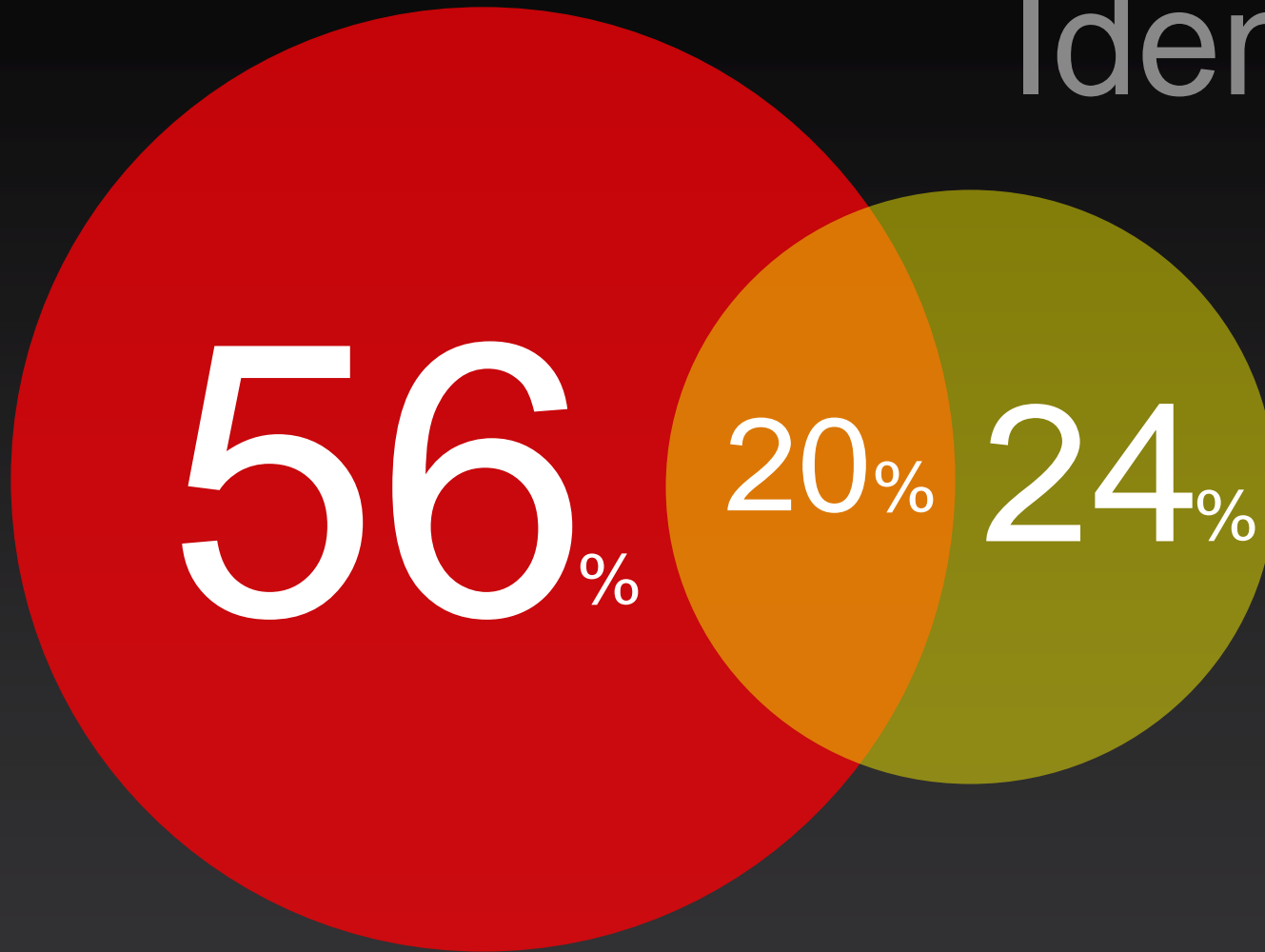
contribute to two or more **single-author** blogs

Bloggers

47%

contribute to two or more **collaborative** blogs

Identity

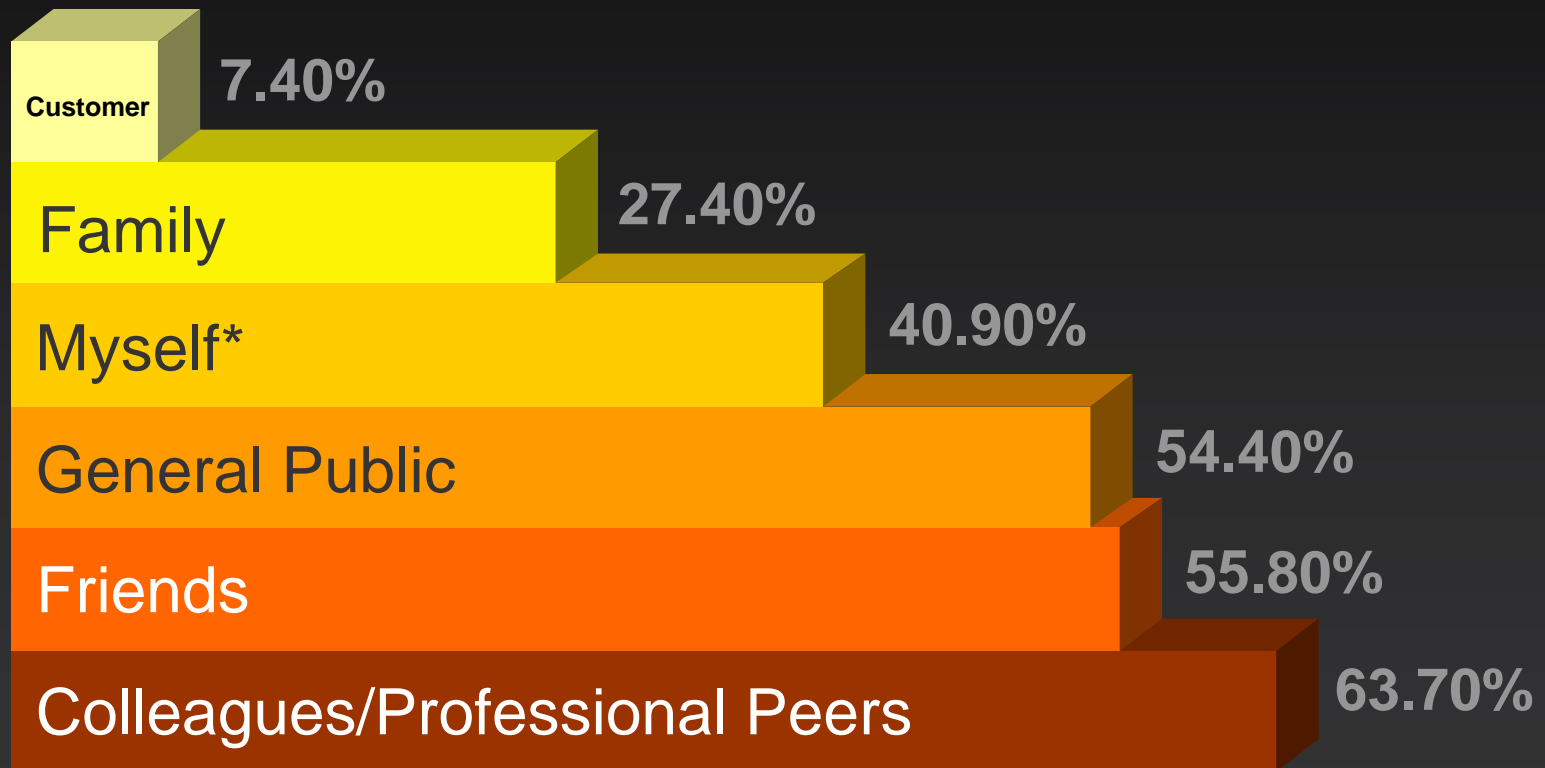


Real name | Both | Pseudonym

Topical Areas

Professional
Technology Personal
Academic News Politics
Hobbies Entertainment Family/friend
Religion Business Health Sport

Audiences



Preservation

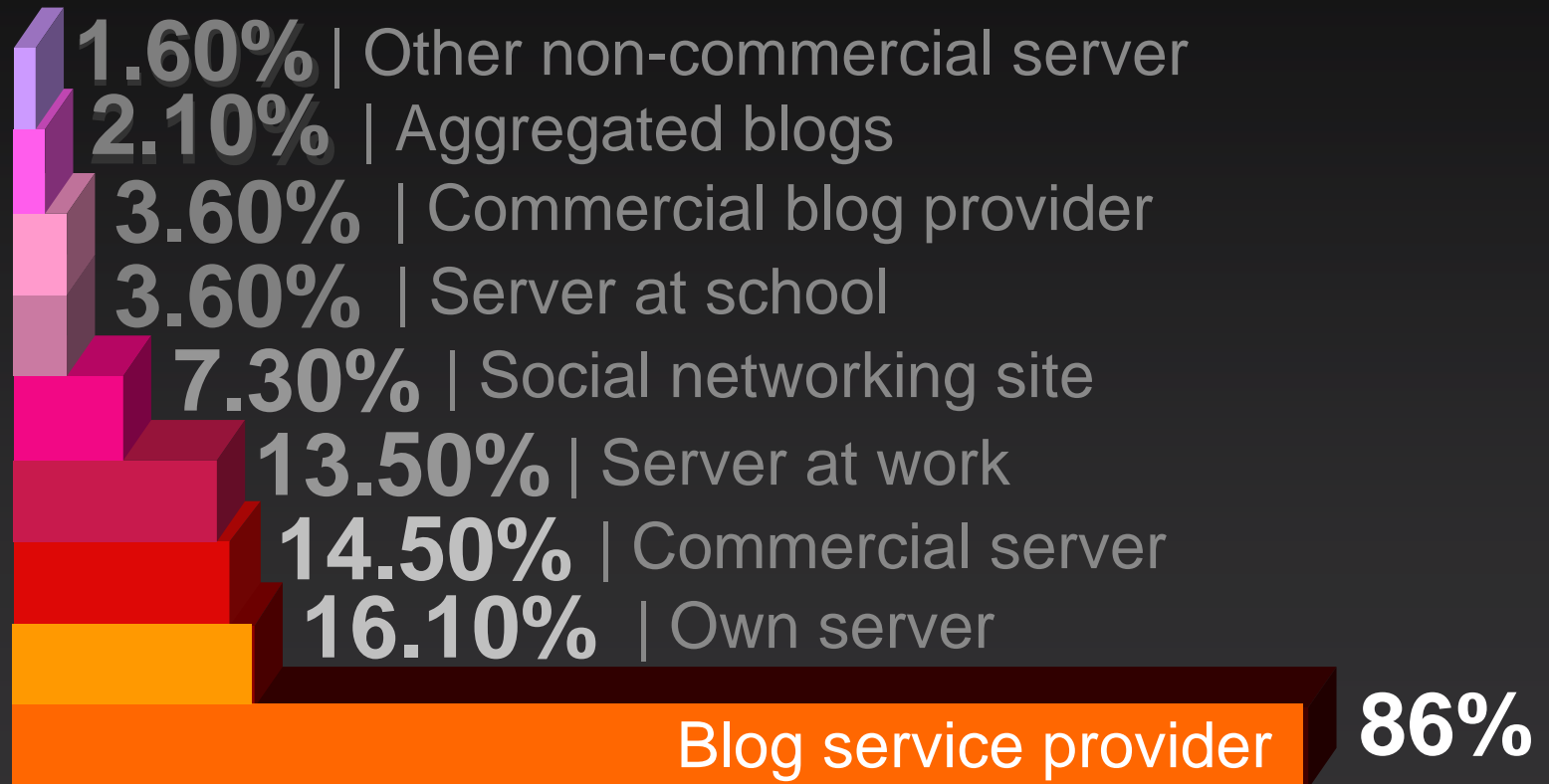
71%

agree that their **own**
blogs should be
preserved

36%

agree that **every**
blog should be
preserved

Hosting



Composing

Often use

blog editor

Alternative way:

Word processor (desktop and online)

Email composer

Mobile phone/PDA

Desktop plug-in blog editor

Instant messenger

Password

2% require
password to
view **ALL**
posts

19%
require password to
view **CERTAIN** posts

EDIT Posts

96%

edited content **after**
it was posted

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

Presentation

84%

have changed the
appearance of their
blogs

Duplication

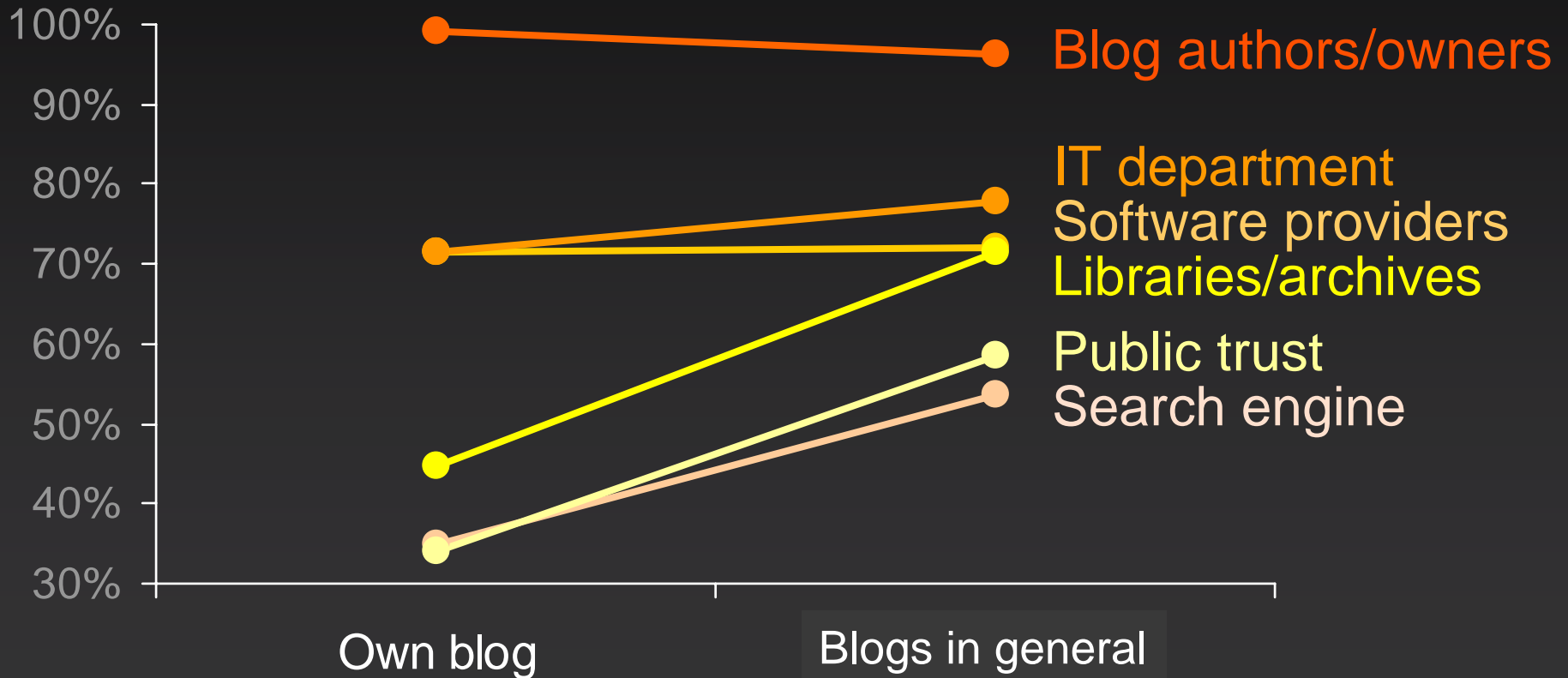
67%

duplicate some **posts**
(18% duplicate all)

36%

duplicate some **comments**
(12% duplicate all)

Responsibility




Characteristics

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

Important

Characteristics

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

Unimportant

Selection Criteria

context

content

style

format

audience

author

Selection Criteria

context

content

style

format

audience

author

Blogosphere

(all, as many)

Representation

(time/society)

Future use

Significance

(cultural, social, political)

Selection Criteria

context

content

style

format

audience

author

Subject

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

Attribute

(original, innovative, quality, authoritative)

Selection Criteria

context

content

style

format

audience

author

Expressive

Citizen journalism

Commentary

Advocacy

Scholarship

Utilitarian

Selection Criteria

context

content

style

format

audience

author

Text Photos

Video Audio Other_multimedia

Comments

Keywords

Links

Imported_images

Bookmarks

Imported_tags Feeds

Selection Criteria

context

content

style

format

audience

author

Popularity

(e.g. readership)

Impact

(e.g. reference, citation)

Location

(e.g. regional, community)

Public discourse

(cultural, social, political)

Audience preference

Expert preference

Selection Criteria

context

content

style

format

audience

author

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

Impact

*“[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.”

Investment

54%

Would not pay to
preserve their **OWN**
blog

ANNUAL

\$1 – 200

Avg. \$36.82

Conclusions

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